The Telecommunications and Data Acquisition Progress Report 42-88

October – December 1986

E. C. Posner
Editor

February 15, 1987

NASA

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-88

October – December 1986

E. C. Posner
Editor

February 15, 1987

NASA

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio astronomy, it reports on activities of the Deep Space Network (DSN) and its associated Ground Communications Facility (GCF) in planning, in supporting research and technology, in implementation, and in operations. Also included is TDA-funded activity at JPL on data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Tracking and Data Systems (OSTDS).

In geodynamics, the publication reports on the application of radio interferometry at microwave frequencies for geodynamic measurements. In the search for extraterrestrial intelligence (SETI), it reports on implementation and operations for searching the microwave spectrum. The latter two programs are performed for NASA's Office of Space Science and Applications (OSSA).

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund which involve the TDA Office are included.

This and each succeeding issue of the TDA Progress Report will present material in some, but not necessarily all, of the following categories:

OSTDS Tasks:

    DSN Advanced Systems
        Tracking and Ground-Based Navigation
        Communications, Spacecraft-Ground
        Station Control and System Technology
        Network Data Processing and Productivity
    DSN Systems Implementation
        Capabilities for Existing Projects
        Capabilities for New Projects
        New Initiatives
        Network Upgrade and Sustaining
    DSN Operations
        Network Operations and Operations Support
        Mission Interface and Support
        TDA Program Management and Analysis
    GCF Implementation and Operations
    Data and Information Systems

OSSA Tasks:

    Search for Extraterrestrial Intelligence
    Geodynamics
        Geodetic Instrument Development
        Geodynamic Science

Discretionary Funded Tasks

Contents

OSTDS TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

COMMUNICATIONS, SPACECRAFT-GROUND

STATION CONTROL AND SYSTEM TECHNOLOGY

DSN Systems Implementation
CAPABILITIES FOR EXISTING PROJECTS

Ka-BAND CAPABILITIES

NETWORK UPGRADE AND SUSTAINING

DSN Operations
MISSION INTERFACE AND SUPPORT

Very Long Baseline Interferometry Using a Radio Telescope in Earth Orbit

J. S. Ulvestad, C. D. Edwards, and R. P. Linfield

Tracking Systems and Applications Section

Successful Very Long Baseline Interferometry (VLBI) observations at 2.3 GHz have been made using an antenna aboard an earth-orbiting spacecraft as one of the receiving telescopes. These observations employed the first deployed satellite (TDRSE-E for East) of the NASA Tracking and Data Relay Satellite System (TDRSS). Fringes were found for 3 radio sources on baselines between TDRSE and telescopes in Australia and Japan. This article describes the purpose of the experiment and the characteristics of the spacecraft that are related to the VLBI observations. It goes on to explore the technical obstacles to maintaining phase coherence between the orbiting antenna and the ground stations, as well as the calibration schemes for the communications link between TDRSE and its ground station at White Sands, New Mexico. System coherence results and scientific results for the radio source observations are presented. Using all available calibrations, a coherence of 84% over 700 seconds was achieved for baselines to the orbiting telescope.

I. Introduction

Very Long Baseline Interferometry (VLBI) is a technique that uses the correlation of data recorded at two or more separate telescopes in order to achieve high angular resolution. For ground-based VLBI, the maximum projected baseline and consequent resolution are limited by the physical diameter of the earth. It has been proposed that a spacecraft dedicated to VLBI be launched so that the orbiting observatory could be used for improved resolution and mapping of compact radio sources. The two major projects currently being studied are the QUASAT project, involving NASA and the European Space Agency, and the RADIOASTRON project under way in the Soviet Union. Reference 1 describes various space VLBI concepts and the scientific benefits of doing

VLBI from space. More up-to-date status reports on QUASAT and RADIOASTRON are given in Refs. 2 and 3.

Several technical hurdles must be cleared in order to use a free-flying spacecraft as a VLBI telescope. First, a local oscillator purity of 10^{-13}–10^{-12} is needed on board the spacecraft. This might be achieved either by having an accurate oscillator in space or by successfully transferring a frequency reference from the ground to the spacecraft via a communications uplink. Second, the spacecraft orbit must be adequately known, and variations in that orbit modelled or calibrated accurately enough to permit lengthy coherence times. Third, the wide-band VLBI data must be transmitted to the ground and recorded there. A demonstration that these problems

could be solved in a VLBI experiment using an element of TDRSS was proposed several years ago (Ref. 4) and carried out in July and August of 1986. A secondary purpose of such an experiment was the investigation of compact extragalactic radio sources on baselines longer than the earth's diameter and the demonstration that a dedicated VLBI observatory in space could accomplish scientifically valuable goals. This article describes the results of the successful space VLBI experiment carried out in 1986. The primary results of the experiment were first reported in Ref. 5.

II. Design of an Experiment Using TDRSE

The TDRSE satellite is in geosynchronous orbit at 41° west longitude, above the western part of the Atlantic Ocean. It has two 4.9 m antennas that operate near 2 GHz (S-band) and 15 GHz (K-band). In addition, there is a smaller antenna that is used for the uplink (15 GHz) and downlink (14 GHz) to the ground control station at White Sands, New Mexico. Each 4.9 m antenna has a 3 dB bandwidth of 16 MHz and a system temperature of approximately 320 K (see Ref. 5 for more details of the antenna calibration). The two 4.9 m antennas have separate intermediate frequency chains that are driven by a common local oscillator on the spacecraft; that oscillator is tied to a cesium frequency standard at White Sands which is transferred to the spacecraft via the 15 GHz uplink. Orbit determination for TDRSE is normally done using range and Doppler data acquired with two widely spaced ground transponders in the Bilateration Ranging Transponder System (BRTS).

The TDRSS system is designed primarily to track and relay communications from satellites in low earth orbit. The normally allowed pointing window for a 4.9 m antenna on the spacecraft is a square 28° on a side, centered at the nadir. This is adequate for covering satellites near the earth, which has an angular radius of about 9° as viewed from geosynchronous orbit. In fact, our observations were further limited by the restriction that we were only allowed to observe sources at southern declinations. A pointing window consisting of the southern half of an ellipse with a semi-major axis of 31° (north-south) and a semi-minor axis of 23° (east-west) was supposed to be available for the VLBI experiment. However, problems with the software used to point the spacecraft antennas prevented us from making any successful observations outside the southern half of the 28° square. Figure 1 displays the field of view of a TDRSE antenna.

The low sensitivity of the spacecraft antenna and the pointing limits restricted the choice of ground antennas that could be used in the orbiting VLBI (OVLBI) experiment. Specifically, very sensitive telescopes on the side of the earth opposite TDRSE were required. Therefore, the observations were made

using the 64 m antenna of the NASA Deep Space Station at Tidbinbilla, Australia; and the 64 m antenna of the Japanese Institute for Space and Astronautical Science (ISAS), at Usuda, Japan. In addition, the 26 m antenna of the Radio Research Laboratory (RRL) at Kashima, Japan was used during most of the OVLBI observations in order to check the performance of the larger ground antennas. After an S-band maser receiver from the Jet Propulsion Laboratory was installed on the telescope at Usuda, both 64 m telescopes had system temperatures of about 15–20 K.

The low spacecraft sensitivity also meant that it was necessary to use the widest feasible bandwidth and the longest possible integrations in the VLBI observations. The large bandwidth was achieved by using the Mark III VLBI recording system (Ref. 6) in Mode E. Seven contiguous 2 MHz channels centered at 2277.99 MHz were recorded, giving a total bandwidth of 14 MHz. This is nearly the maximum bandwidth available from TDRSE. The data from the celestial radio sources were recorded directly in the usual way at Usuda and Tidbinbilla. The wide-band data received by TDRSE were coherently translated to the downlink frequency of 14 GHz, transmitted to the ground, translated to an intermediate frequency of 370 MHz, and fed directly into the video converters of the Mark III recorder. Details of the methods used to achieve the greatest possible system coherence (hence, the longest possible integration times) are given in Section III below.

III. Coherence and Stability

A. Coherence Requirements for Interferometry

The coherence for an integration time T can be defined:

$$C(T) = \left\| \frac{1}{T} \int_0^T dt \, e^{i\widetilde{\phi}(t)} \right\|$$

where $\widetilde{\phi}$ is the residual interferometer phase after removal of a linear trend (Ref. 7). For $\widetilde{\phi}(t) \ll 1$ radian, $C(T) \approx 1$; as the phase residuals grow to an order of 1 radian or more, the coherence drops off quickly. Since the observed fringe amplitude is proportional to the coherence, $C(T)$ can be thought of as an efficiency factor.

Based on this dependence of the coherence on the rms phase fluctuations, one can formulate a rough stability requirement for maintaining coherence. For an integration time of length τ at an observation frequency ν_{RF}, the Allan (or two-point) variance $\sigma_y(\tau)$ must satisfy:

$$\nu_{RF} \, \tau \, \sigma_y(\tau) < 1 \text{ radian}$$

The minimum value of τ is given by signal-to-noise considerations. To have a high probability of fringe detection on the baseline between a TDRSE 4.9 m antenna and either of the 64 m ground antennas, it was necessary to ensure good coherence out to 500 s. Given an observing frequency of 2.3 GHz, this translates to a stability requirement of $\sigma_y(\tau) < 1 \times 10^{-12}$.

B. Contributions to Phase Instability of the Experiment

Some of the sources of phase instability for the OVLBI experiment were typical of any VLBI experiment; others were a direct consequence of using an orbiting antenna. First consider spacecraft-specific error sources. The most important consideration here is the effect of spacecraft motion. Changes in the spacecraft position change the geometric delay between TDRSE and the ground stations. The observation frequency of 2.3 GHz corresponds to a wavelength of about 13 cm. Thus a 13 cm change in spacecraft position (in the direction of the observed quasar) would change the geometric phase by a full cycle. Even more important, however, is the effect of spacecraft motion on the communications link between TDRSE and the White Sands ground station. The local oscillators onboard TDRSE are derived from a 15 GHz pilot tone uplink from White Sands, while the observed signal is transmitted back to White Sands on a 14 GHz downlink. After accounting for frequency translation onboard the spacecraft, the relevant frequency at which spacecraft position errors affect the TDRSE-White Sands communications link is roughly twice the downlink frequency, or more than 25 GHz, which corresponds to an effective wavelength of about 1 cm. Namely, a 1 cm change in the line-of-sight distance between TDRSE and White Sands will result in nearly a full cycle phase shift of the signal recorded at White Sands. To maintain good coherence for integration times of 500 s, it follows that TDRSE accelerations must be modeled or calibrated to better than 10^{-8} m/s^2.

Other error sources are more typical of ground-based observations. Independent frequency standards are used at each site to generate local oscillators and to time tag the recorded data. A hydrogen maser, with $\sigma_y(\tau = 500 \text{ s}) < 1 \times 10^{-14}$, was used at the DSN 64 m station in Australia. Usuda used a rubidium standard, with $\sigma_y(\tau = 500 \text{ s}) \approx 2 \times 10^{-13}$. At White Sands, a cesium standard with stability $\sigma_y(\tau = 500 \text{ s}) \approx 4 \times 10^{-13}$ was used as the frequency reference for all the TDRSS local oscillators, both or the ground and onboard the spacecraft, via the uplink pilot tone. A hydrogen maser was used, however, to drive the Mark III terminal equipment at White Sands.

Propagation media effects also contribute to phase instability. The troposphere will affect the 2.3 GHz observations at the ground antennas, but more important in terms of phase stability is its effect on the higher frequency TDRSE-White Sands links. For the ionosphere the situation is reversed, as the phase effect of the ionosphere scales as $1/\nu$. The ionosphere should not be a factor for the uplink and downlink to TDRSE, due to the high link frequencies, but will be important at the 2.3 GHz observation frequency.

Finally, instrumental instabilities must be considered. Of particular concern is the stability of the TDRSE spacecraft instrumentation. This includes both the physical stability of the 4.9 m antennas and the stability of the various RF receivers, mixers, and local oscillators.

C. December 1985 Stability Test Results

To determine the suitability of TDRSS for OVLBI, tests were performed at White Sands in December, 1985. In one configuration, a 2.3 GHz beacon at White Sands was observed by one of the 4.9 m TDRSE antennas and transferred back to White Sands, where the tone was tracked and phase residuals recorded. Later, the TDRSE ephemeris based on BRTS data (standard range and doppler measurements) was used to remove the effects of spacecraft motion. Phase fluctuations of roughly ±1/4 cycle remained, with coherence falling quickly from 95% at 100 s to less than 60% at 500 s.

To improve coherence, a second test was performed in which both 4.9 m antennas observed the ground beacon. Phase residuals observed in one channel were used to correct the phase in the other channel, thereby calibrating all effects on the TDRSE-White Sands link common to both channels, including propagation media, common spacecraft instrumental instabilities, spacecraft motion, and instabilities in the ground frequency standard. In this configuration, the phases observed in the two channels tracked each other to about ±0.02 cycles, yielding a relative coherence of better than 99% for 1000 s.

D. TDRSE Ephemeris Calibrations

Based on the December stability tests, it was decided that the astronomical observations would employ one TDRSE 4.9 m antenna to track a 2.3 GHz ground beacon while the other 4.9 m antenna observed an extragalactic radio source. This setup provided two sources of information on spacecraft position: the BRTS data and the tracked beacon phase. The BRTS data provided a full three-dimensional spacecraft ephemeris, while the beacon data yielded more accurate information, but only along the TDRSE-White Sands line-of-sight. Figure 2 depicts the configuration used for the quasar observations, while Fig. 3 is a simplified block diagram of the overall system.

During correlation at Haystack Observatory, delay and phase models must be supplied which account for the effects

of TDRSE motion. The model interferometer delay for the baseline between TDRSE and Usuda (for example) is given by:

$$c\tau = [\vec{r}_{USUDA} - \vec{r}_{TDRSE}] \cdot \hat{\epsilon}_{QSO} + \|\vec{r}_{WS} - \vec{r}_{TDRSE}\|$$

where $\vec{r}_{TDRSE}$ and $\vec{r}_{USUDA}$ are the position vectors for TDRSE and Usuda, $\vec{r}_{WS}$ is the position vector for the White Sands ground station, and $\hat{\epsilon}_{QSO}$ is a unit vector in the quasar direction. This delay is a sum of two terms: the first is the standard geometric interferometer delay representing the difference in arrival times at the two antennas for a wavefront from the quasar, while the second is the one-way propagation delay between the TDRSE antenna, where the wavefront is observed, and the White Sands ground station, where the signal is recorded and time-tagged.

For ground-based VLBI, the phase model is simply the delay model multiplied by the observation frequency. Due to the K-band communication links between TDRSE and White Sands, however, the situation is more complicated for this experiment. The phase model takes the form:

$$\phi = \frac{[\vec{r}_{USUDA} - \vec{r}_{TDRSE}] \cdot \hat{\epsilon}_{QSO}}{c} \nu_{obs}^{QSO} (S)$$

$$+ \frac{\|\vec{r}_{WS} - \vec{r}_{TDRSE}\|}{c} \left(2\nu_{down}^{QSO} (K) - \nu_{obs}^{QSO} (S)\right)$$

where $\nu_{obs}^{QSO} (S)$ is the 2.3 GHz quasar observation frequency and $\nu_{down}^{QSO} (K)$ is the 14 GHz downlink frequency for the quasar data. Again the model is the sum of two terms: a geometric interferometer phase and a link term. The first term, representing the geometric phase, is indeed simply the interferometer delay multiplied by the 2.3 GHz S-band observation frequency. The link term, however, consists of the one-way link delay scaled by a frequency of roughly twice the K-band downlink frequency of 14 GHz. The exact value of the frequency factor can be understood as follows: The uplink pilot tone is used to generate a local oscillator frequency onboard the spacecraft of $\nu_{down}^{QSO} (K) - \nu_{obs}^{QSO} (S)$, which is subsequently used to frequency translate the observed quasar signal to the downlink frequency $\nu_{down}^{QSO} (K)$. Any delay errors on the uplink will therefore affect the phase at the effective frequency of $\nu_{down}^{QSO} (K) - \nu_{obs}^{QSO} (S)$. The relevant frequency for the downlink is $\nu_{down}^{QSO} (K)$, and so the total effective frequency for the TDRSE link is $2\nu_{down}^{QSO} (K) - \nu_{obs}^{QSO} (S)$. Because of this very high effective frequency (of order 25 GHz), the link term for the model phase is more sensitive to spacecraft position errors than the geometric term.

The phase of the White Sands beacon signal provides a highly accurate measurement of the link term. The round trip phase for the beacon signal is

$$\phi_{beacon} = \frac{\|\vec{r}_{TDRSE} - \vec{r}_{WS}\|}{c} \left(2\nu_{down}^{beacon} (K)\right)$$

where $\nu_{down}^{beacon} (K)$ is the 14 GHz downlink frequency for the beacon data. As the accuracy requirements for the delay model are not stringent, BRTS data were used to calculate the entire delay model. The BRTS data were also used to calculate the geometric term in the phase model, since a full three dimensional TDRSE position vector is required. The beacon data, however, were used to calibrate the link term in the phase model. Details of how the individual corrections were applied at the correlator are described in Section IV.

In addition to accounting for the effects of spacecraft motion, the beacon calibration also removes any effect of the troposphere on the link, since both error sources have the same signature in the link phase. In fact, any non-dispersive error source which corrupts the link will be completely eliminated by the beacon calibration. Finally, spacecraft instrumental phase instabilities common to both channels are largely removed.

E. Calibration of the White Sands Cesium Standard

A hydrogen maser was used as a frequency reference for the Mark III terminal at White Sands. All of the TDRSS local oscillators, however, were by necessity referenced to the TDRSS cesium standard. By calibrating this cesium standard against the hydrogen maser, it was possible to remove the effects of the cesium instability. A comb generator was used to generate a 370 MHz harmonic of the cesium 5 MHz reference. This tone was then mixed down to baseband (0-2 MHz) with a local oscillator derived from the hydrogen maser, and tracked in realtime with a special purpose tone tracker (also referenced to the hydrogen maser) which extracted residual phases at 2 s intervals for the tracked tone. These residuals were used during fringe fitting at the Haystack Observatory correlator in order to correct for the instabilities of the TDRSS cesium.

IV. Data Correlation Procedure

With VLBI data on baselines between earth telescopes, the correlator delay and phase models are fairly simple. The total delay is the geometric delay — the difference in arrival time at the two antennas. The phase is the product of this delay and the sky (observing) frequency. The delay and phase at any epoch can be calculated from only a few parameters: the three

components of the vector baseline in an earth-fixed frame, the longitude of the baseline, and the radio source hour angle and declination. The time variation of delay and phase is given by the rotation of the earth (uniform solid-body rotation is a sufficiently accurate approximation for correlation). Various additional corrections are needed for astrometric or geodetic measurements, but not for measurements of correlated fluxes.

Data from this experiment were correlated with the Haystack Observatory Mk IIIA correlator. The delay and phase models were entered in several stages. The first order models were derived from a constant (earth-fixed) position for the spacecraft (TDRSE is in geosynchronous orbit). This position was obtained from the spacecraft ephemeris, using an epoch near the middle of the observation to be correlated. The approximate geometric delay and phase from the first order models were obtained in the same manner as for two earth-based antennas.

During correlation, the second-order models were applied. Correction polynomials in delay and phase supplied the remainder of the information from the orbit ephemeris (i.e., that not present in the first-order model). These polynomials corrected the geometric delay and phase, and specified the link delay, to acceptable accuracy.

Third order corrections were applied during fringe-fitting. A table of phase residuals was provided, one value for each 2 s correlator output record. The correlated data were rotated by these residuals before being Fourier-transformed. The phase residuals applied in this process arose from measurements made at White Sands (see Sections IIID and IIIE). There were two distinct terms contributing to these residuals. The first was the sum of the high frequency terms in the link phase measurements (i.e., the residuals remaining after a cubic polynomial fit). The second contribution was the measured phase difference between the hydrogen maser and the cesium frequency standard at White Sands.

V. Results

A. Coherence Measurements

Coherence measurements derived from the correlated data are shown in Fig. 4. The lowest curve (A) is the coherence achieved without applying phase residuals (the third-order correction described above) during fringe-fitting. The middle curve (B) shows the improvement that is achieved when the measured phase differences between the cesium and hydrogen maser frequency standards are applied. As the figure shows, the improvement is small at short time scales and large at long time scales. This is consistent with experience with cesium

standards in VLBI experiments — good coherence for short time scales and poor coherence at long time scales. The upper curve (C) shows the coherence when the phase residuals for both the link calibration and frequency standards are applied. The link calibration corrects most of the short-term coherence loss. It is unclear why this loss occurred. It may be due to physical motion of the spacecraft, with amplitudes of a few mm and time scales of 10–100 s. Alternatively, it may be due to instabilities in the spacecraft electronics which are common to the two channels (quasar and beacon).

When all corrections were applied, the coherence was 92% at 300 s and 84% at 700 s (Fig. 4, curve C). If the beacon phase is not used, and instead the BRTS ephemeris is used to calibrate the link phase, the coherence falls to 84% at 300 s and 78% at 700 s (Fig. 4, curve B). Finally, if the phase residuals for the cesium clock also are not used (Fig. 4, curve A), the coherence drops to 73% at 300 s and 60% at 700 s. Given that the high link frequency was more important than the observing frequency in limiting the coherence, and given the fact that TDRSE was not designed for VLBI, it is likely that the coherence would be nearly as good (or perhaps better) for a dedicated VLBI observatory observing at frequencies as high as 20–30 GHz.

B. Astronomical Results

The telescopes in Australia and Japan were calibrated in the normal way. A noise diode was used to monitor the system temperature, and observations of sources with known flux were used to measure antenna gain. The TDRSE spacecraft was equipped with an automatic gain control, so that a different calibration method was used, as described in Ref. 5. The basic procedure involved the knowledge that the projected baseline between TDRSE and Australia was sometimes similar to projected baselines on ground baselines at other times. Because the correlated flux on these baselines was known from the calibration of the ground antennas, the sensitivity of TDRSE could be determined. The result was that TDRSE was determined to have an aperture efficiency of 0.4 and a system temperature of 320 K. By comparison, the ratio of gain to system temperature for the 64 m ground antennas was 4,000–5,000 times that for TDRSE.

Correlated flux was detected on a total of six scans for three sources: 1510-089, NRAO 530 (1730-130), and 1741-038, all quasars. All scans for which correlated flux was not detected were known to have had equipment setup or antenna pointing errors. In other words, none of the nondetections were due to inadequate correlated flux on long projected baselines. The maximum baseline was 1.4 earth diameters, on NRAO 530, with baselines longer than one earth diameter obtained for all three sources. The sampling of projected baselines in the u-v plane for NRAO 530, the source for which

the most data were obtained, is shown in Fig. 5. Here, u is the east-west component of the baseline, while v is the north-south component. The track at $v = 1.1 \times 10^8\lambda$ was generated by the TDRSE-Japan baseline, and the track at $v = 5.5 \times 10^7\lambda$ was generated by the TDRSE-Australia baseline. The point at $u = 10^7$, $v = 5.5 \times 10^7\lambda$ was generated by the Japan-Australia baseline. The u-v coverage shown in Fig. 5 is too poor to make a map of a radio source.

Fringe visibilities on the maximum projected baselines were 0.15 (1510-089), 0.05 (NRAO 530), and 0.50 (1741-038). A plot of the correlated flux as a function of projected baseline length is shown in Fig. 6 for 1510-089 and NRAO 530, the two sources with the most data. The correlated flux density drops rapidly with increasing baseline length for NRAO 530, but appears to flatten at 0.3 Jy, implying a brightness temperature of $\approx 4 \times 10^{11}$ K. The correlated flux of 1510-089 drops off more slowly with baseline length. For both sources, even longer baselines are needed to resolve the innermost structure.

VI. Summary

The coherence results described in this article demonstrate the feasibility of transferring a coherent frequency reference to an orbiting satellite with sufficient stability to perform VLBI using a telescope in earth orbit. Careful calibration of the spacecraft communications link is required to achieve the highest coherence. It has also been shown that quasars still contain significant amounts of compact flux on baselines longer than an earth diameter. Togther, these two results demonstrate the technical and scientific feasibility of doing VLBI using a dedicated radio telescope in earth orbit.

A second set of VLBI observations involving the TDRSE satellite is scheduled for January 1987. One of the goals of those observations will be to better understand the limiting error sources after all calibrations have been applied. Debugging of the pointing software for the spacecraft antennas has been accomplished, so the larger pointing window described in Section II will be usable. The increased flexibility in pointing will be used to observe sources at more southerly declinations and at greater eastern and western elongations than was possible in the first experiment. Such observations will achieve projected baselines of 2-2.5 earth diameters (25,000-30,000 km) on some radio sources. Such baselines will extend the investigations of source structure further beyond the limits imposed by the earth's physical size and will provide a probe of very compact radio emission and high brightness temperatures.

References

1. European Space Agency, "Quasat — a VLBI observatory in space," Proceedings of a Workshop held at Gross Enzersdorf, Austria, 18-22 June 1984.

2. Jordan, J. F., "Quasat — U.S. Status Report," in *Proceedings of the Green Bank Workshop on Radio Astronomy from Space*, National Radio Astronomy Observatory, Green Bank, West Virginia, September 30-October 2, 1986, in press.

3. Schilizzi, R. T., "The U.S.S.R. RadioAstron Space VLBI Project," in *Proceedings of the Green Bank Workshop on Radio Astronomy from Space*, National Radio Astronomy Observatory, Green Bank, West Virginia, September 30-October 2, 1986, in press.

4. Levy, G. S., Christensen, C. S., Jordan, J. F., Preston, R. A., and Burke, B. F., "Orbiting Very Long Baseline Interferometer Demonstration using the Tracking and Data Relay Satellite System," in *Proceedings of International Astronomical Union Symposium No. 110, VLBI and Compact Radio Sources*, Bologna, Italy, June 27-July 1, 1983, edited by R. Fanti, K. Kellermann, and G. Setti, pp. 405-406.

5. Levy, G. S., Linfield, R. P., Ulvestad, J. S., Edwards, C. D., Jordan, J. F., Di Nardo, S. J., Christensen, C. S., Preston, R. A., Skjerve, L. J., Stavert, L. R., Burke, B. F., Whitney, A. R., Cappallo, R. J., Rogers, A. E. E., Blaney, K. B., Maher, M. J., Ottenhoff, C. H., Jauncey, D. L., Peters, W. L., Nishimura, T., Hayashi, T., Takano, T., Yamada, T., Hirabayashi, H., Morimoto, M., Inoue, M., Shiomi, T., Kawaguchi, N., and Kunimori, H., "Very Long Baseline Interferometric Observations Made with an Orbiting Radio Telescope," *Science*, Vol. 234, pp. 187-189, October 10, 1986.

6. Rogers, A. E. E., Cappallo, R. J., Hinteregger, H. F., Levine, J. I., Nesman, E. F., Webber, J. C., Whitney, A. R., Clark, T. A., Ma, C., Ryan, J., Corey, B. E., Counselman, C. C., Herring, T. A., Shapiro, I. I., Knight, C. A., Shaffer, D. B., Vandenberg, N. R., Lacasse, R., Mauzy, R., Rayhrer, B., Schupler, B. R., and Pigg, J. C., "Very-Long-Baseline Radio Interferometry: The Mark III System for Geodesy, Astrometry, and Aperture Synthesis," *Science*, Vol. 219, pp. 51-54, January 7, 1983.

7. Rogers, A. E. E., and Moran, J. M., "Coherence Limits for very long baseline interferometry," *IEEE Transactions on Instrumental Measurements*, Vol. IM-30, No. 4, pp. 283-286, 1981.
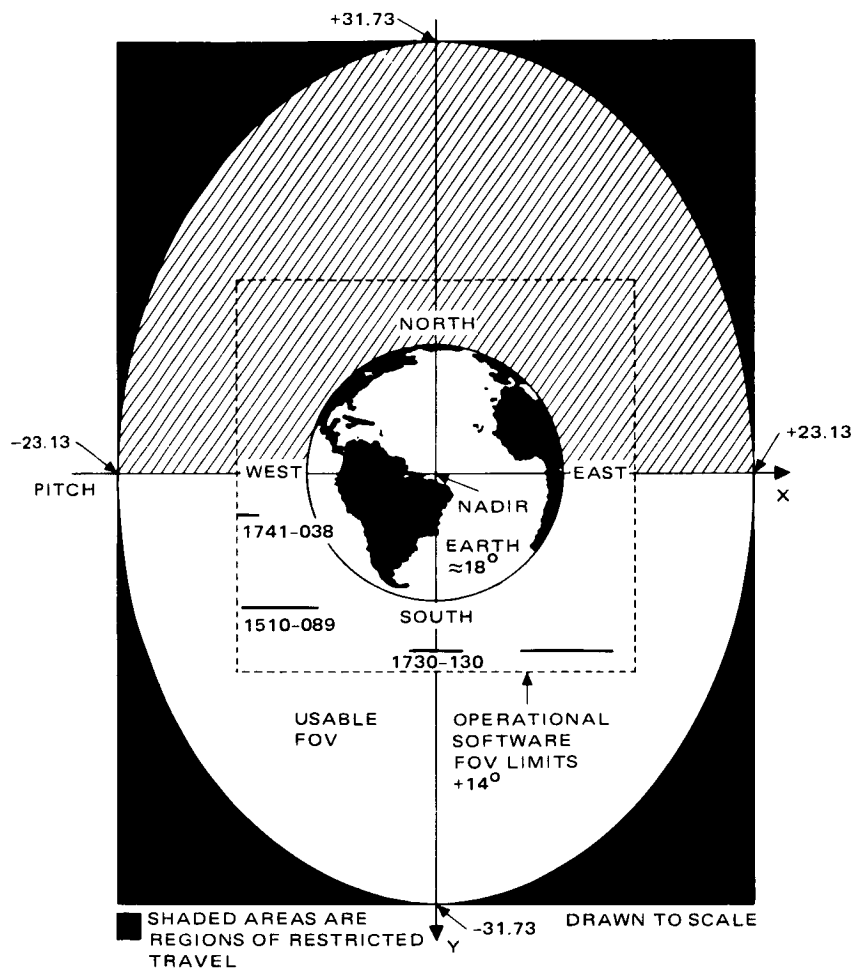
Fig. 1. Field of view of the TDRSE antenna. The large ellipse is the large pointing window specified for the antenna, while the smaller square is the field-of-view (FOV) normally allowed by the operational software. Observations in the hatched region were not allowed. Approximate locations of the radio sources successfully observed are indicated.
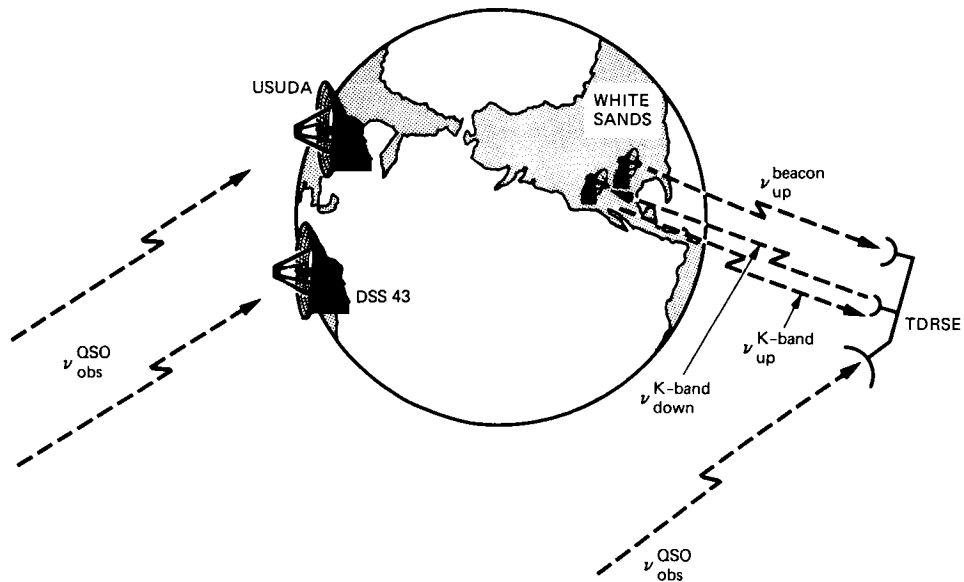
Fig. 2. Link configuration for the TDRSS OVLBI observations
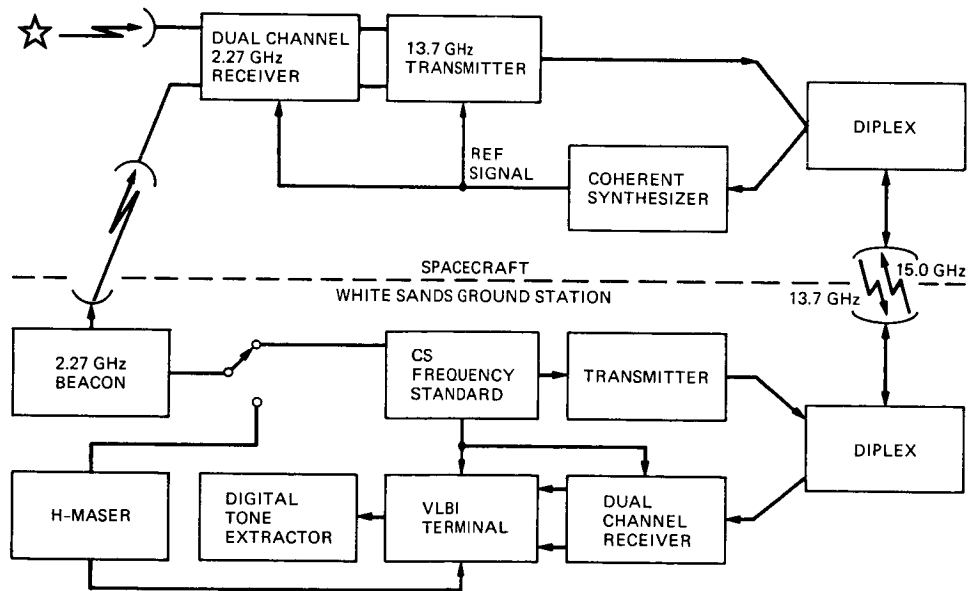
Fig. 3. Simplified block diagram of the configuration used for VLBI obserevations made with the TDRSE spacecraft
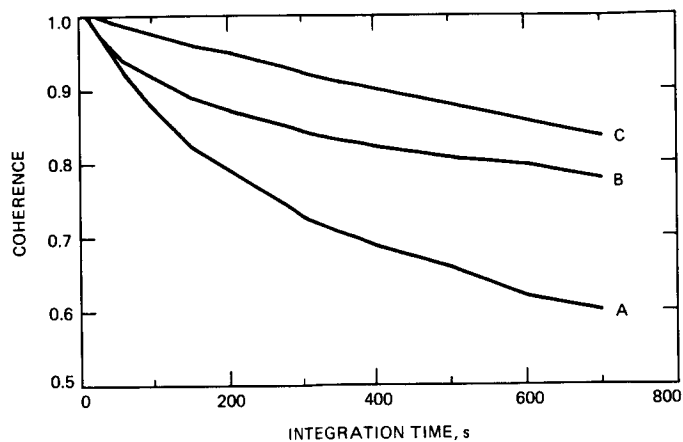
Fig. 4. TDRSS OVLBI coherence results for three different cases. Corrections included in generating curves A, B, and C are described in text.
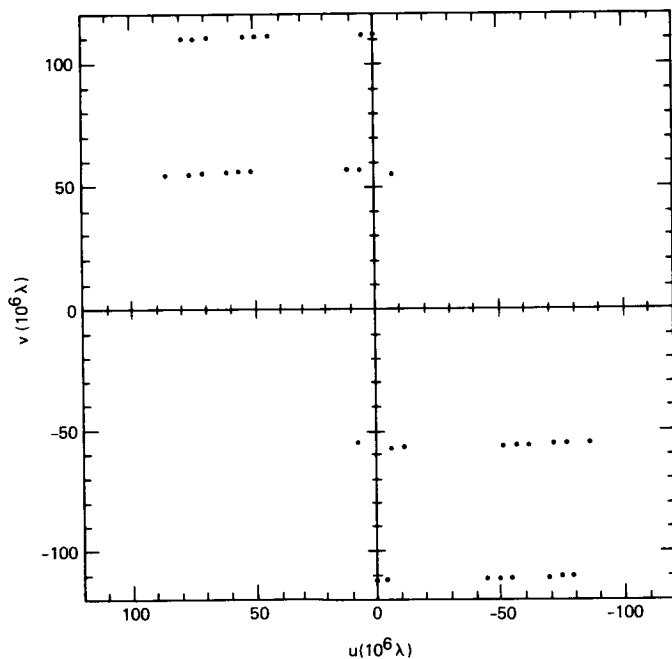
Fig. 5. Coverage of the $u-v$ plane for the observations of NRAO 530. The east−west projection is specified by u and the north−south projection is given by v, both in units of millions of S-band wavelengths.
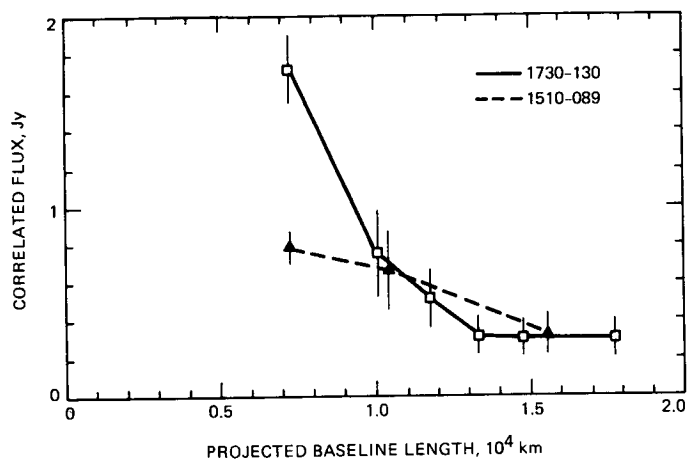
Fig. 6. Plots of correlated flux versus projected baseline length ("visibility plots") for observations of two quasars, as found in the TDRSS OVLBI experiment. The earth diameter is approximately 1.3×10^4 km.

N87-17945

Directional Couplers for Detecting the TE_{11} and TE_{12} Circular Waveguide Modes

D. Hoppe

Radio Frequency and Microwave Subsystems Section

This article describes the theoretical and experimental results for a pair of mode-selective directional couplers designed to detect the TE_{11} and TE_{12} modes in a multimode circular waveguide. A brief description of the design of the couplers is presented, followed by a comparison of their measured and calculated parameters. The couplers were used to measure the characteristics of a circular waveguide mode converter. The results of these measurements are described.

I. Introduction

This article is the last in a series describing the transmission-line components required for a proposed high-power (400-kW) millimeter-wave (34.5-GHz) transmitter. The high-power and high-frequency requirements pose many problems that are unique to this transmitter (Ref. 1). In order to transmit power from the amplifier to the antenna focus without breakdown, an oversized multimode circular waveguide must be used. Directional couplers that are sensitive to individual modes are required for monitoring transmitter output power, and for testing the waveguide components and the proposed gyroklystron amplifier. This article describes theoretical and experimental results for two such couplers, one for detecting the TE_{11} circular waveguide mode, and the other for detecting the TE_{12} mode.

II. Mode-Selective Coupler Design

The theory used to design the mode-selective directional couplers is basically the loose coupling theory described by

Miller (Ref. 2). The main circular waveguide is coupled to a rectangular waveguide through a series of coupling holes as depicted in Fig. 1. For TE couplers, the coupling is designed to be through the longitudinal magnetic field (H_z) in the circular and rectangular waveguides; therefore, the rectangular guide is attached to the circular guide through its small (b) dimension as shown in Fig. 1. This configuration eliminates any coupling of TM power from the main waveguide.

Coupling to a particular TE mode in the circular guide is maximized by adjusting the rectangular waveguide's wide (a) dimension so that the phase velocity of its dominant mode matches that of the circular waveguide mode of interest. The coupling waveguide is then tapered down to a standard size for connection of a detector. High directivity is assured by spacing the holes at one quarter of a guide wavelength. Since the coupler is mode-selective, it must by definition reject power present in other circular waveguide modes traveling in the forward and reverse directions. This rejection is usually specified in dB and is called the selectivity. Selectivity for specified modes can be optimized by using a large number of

coupling holes and tapering the coupling profile with distance (i.e., changing hole size). These effects can be simulated quite easily with a simple computer program that adds up the contributions from each hole in the coupling array to obtain the total coupling for each circular waveguide mode at each frequency of interest. Finite wall thickness and the effects of various coupling profiles can be included in the model with little complication.

III. Final Design Parameters

TE_{11} and TE_{12} couplers were designed for a circular waveguide with a diameter of 1.75 in. operating at a center frequency of 34.5 GHz. Design goals included a directivity of at least 40 dB and a selectivity of at least 30 dB for each of the forward and reverse traveling spurious TE_{1n} modes over a 1% bandwidth.

The TE_{11} coupler requires a highly oversized coupling guide with a wide dimension of 1.493 in. in order to match the phase velocities of the TE_{11} circular mode and the TE_{10} rectangular mode. Since for the frequencies of interest and these dimensions the TE_{30} mode is allowed to propagate in the coupling guide (and it is strongly coupled to the TE_{12} mode in the circular guide), great care must be taken in tapering the coupling guide down to standard WR28 size. A highly nonlinear taper was used to avoid coupling large amounts of this TE_{30} power into the TE_{10} mode during the transition. A first prototype of the coupler used a linear taper and showed poor selectivity for the TE_{12} mode due to these effects. In both couplers, the small dimension of the coupling guide was chosen to be the standard WR28 dimension of 0.14 in. The number of holes required to meet the selectivity requirements was determined to be 154 through computer simulation, and a linearly tapered coupling profile was also found to be necessary in order to meet the specifications. Hole spacing was 0.087 in. and wall thickness was 0.020 in. The actual coupling value can be set by scaling the hole diameters, and a coupling value of about 70 dB was obtained using a maximum hole diameter of 0.060 in. A photograph of the fabricated TE_{11} coupler is shown in Fig. 2.

The design of the TE_{12} coupler is similar, requiring a coupling guide width dimension of 0.516 in. and 144 coupling holes with a spacing of 0.092 in. Once again, a linearly profiled coupling profile was used. A maximum hole diameter of 0.050 in. gave a coupling value near 55 dB at the design frequency of 34.5 GHz. The predicted coupling and selectivities for the forward traveling spurious modes are displayed for this coupler versus frequency in Fig. 3. Selectivities for the reverse traveling modes are predicted to be greater than 60 dB and are not plotted.

IV. Coupler Results

Measurements were performed to determine the coupling, directivity, and selectivity values for the two couplers using an HP 8510 automatic network analyzer and additional components required to extend the instrument to the 26.5- to 40-GHz range. The measurement of the TE_{11} coupling is compared to the theoretical results in Fig. 4. A WR28-to-circular waveguide transition and taper up to the 1.75-in. circular waveguide size were used to launch a pure TE_{11} mode into the 1.75-in. circular waveguide in order to make the measurement. The loss through these components was calculated and removed from the measured data. The difference between the theory and measurement is within the error specification for the HP 8510 network analyzer for these large values of attenuation. A measurement of the reverse coupling was also attempted, but no reverse coupling was detectable down to the noise floor of the HP 8510, which is about -100 dB. From this, we conclude that the directivity is greater than 30 dB.

A similar measurement was made for the TE_{12} coupler. In this case, a TE_{11} to TE_{12} mode converter was included in the experimental setup. This device, which was described in a previous report (Ref. 3), is capable of generating a highly pure TE_{12} mode over a limited bandwidth around 34.5 GHz. The measured coupling value of 53.49 dB at 34.5 GHz for the TE_{12} coupler may be compared to the calculated value of 55.50 dB. The agreement here is seen to be somewhat worse than that for the TE_{11} coupler, but still satisfactory considering the possible measurement errors and approximations made in the theoretical calculations.

The measured reverse coupling at 34.5 GHz once again approached the limits of the HP 8510, reaching 99.4 dB. This translates to a directivity near 46 dB at the center frequency. Further measurements involving the mode converter and the two couplers are described later in this article.

An effort was also made to determine the selectivities for the couplers. Since with the available mode converter and tapers it was possible to produce only relatively pure TE_{11} and TE_{12} modes, the TE_{11} coupler's rejection of the TE_{12} mode and the TE_{12} coupler's rejection of the TE_{11} mode were measured. A known composition of TE_{11} and TE_{12} modes was injected into each of the couplers. The phase between the two modes entering the couplers was then changed by adding straight sections of guide before the couplers. As the length is changed by at least one beat wavelength, all relative phases between the two modes can be observed. The interference caused by the unwanted mode can then be seen as a ripple in the coupled power. The technique is very similar to the sliding short method of determining coupler directivity. By placing a limit on the observed ripple, and by know-

ing the mode composition in the waveguide, one can calculate the desired selectivity. Measurements at the center frequency of 34.5 GHz revealed that the selectivity of the TE_{11} coupler for the TE_{12} mode was slightly better than 20 dB, while the TE_{12} coupler rejected the TE_{11} mode by more than 30 dB, and thus met the selectivity goal. Although the selectivity of the TE_{11} coupler with respect to the TE_{12} mode was lower than expected, this should not pose a serious problem in the expected applications of the coupler. This is the case because the coupler will be used to detect the TE_{11} mode, where only small amounts of spurious power are expected to be present, whereas the TE_{12} coupler will be used to detect small amounts of TE_{12} power in the presence of large TE_{11} signal levels.

V. Mode Converter Measurements

This section of experimental results discusses measurements of the frequency response and mode content versus distance for the TE_{11} to TE_{12} mode converter. These measurements were also made using the far-field pattern measurement method described previously (Ref. 3). Once mode-selective couplers have been built and characterized, the above measurements can be made more easily and directly. The directional coupler method is also better suited to high-power measurement than the far-field pattern method.

The frequency response of the TE_{11} to TE_{12} mode converter is depicted in Fig. 5. For these measurements, a pure TE_{11} mode enters the mode converter, and the relative power in the TE_{11} and TE_{12} modes exiting the device is measured using the mode-selective couplers. Theoretical and measured results are given for both modes. Near the center frequency of the converter, nearly all of the input power is converted into the TE_{12} mode. Therefore, the power coupled out by the TE_{12} coupler peaks, while the TE_{11}-coupled power dips. Good agreement is found for all frequencies for the TE_{12} power. Good agreement is also found for the TE_{11} power at all frequencies except for those near maximum TE_{12} output and minimum TE_{11} output. Here, only a 20-dB drop in TE_{11} power is seen, even though a much larger drop is predicted. This is to be expected due to the finite selectivity of the TE_{11} coupler with respect to the TE_{12} mode. The TE_{11} coupler only rejects the TE_{12} mode by 20 dB, so when the TE_{12} power exceeds the TE_{11} power by more than 20 dB, the TE_{11} coupler's output power is primarily due to TE_{12}.

leakage. These results are then consistent with the previous measurements, which showed 20-dB selectivity for the TE_{11} coupler.

Next, the mode content as a function of distance along the mode converter at a frequency of 34.5 GHz was measured using the couplers. This is possible since the mode converter, which is a ripple-walled circular waveguide, was manufactured in several sections. The complete device contained 7 ripples, and was made up of two sections with 2 ripples each, and one section with 3 ripples. By using different combinations of these sections, the mode content after 0, 2, 3, 4, 5, and 7 ripples may be determined. The theoretical mode content may be calculated for any position in the device, however, and is plotted in Fig. 6. The mode content after 0, 2, 3, 4, 5 and 7 ripples as measured by the couplers is also plotted for comparison. For reference purposes, each ripple is 1.364-in. long. Once again, excellent agreement is found for almost all of the data points. There is some disagreement for the 0- and 7-ripple cases due to the finite selectivities of the couplers. For the 0-ripple case, essentially no TE_{12} power is present, but the TE_{12} coupler does produce some output power due to its finite selectivity for the TE_{11} mode. Note that the measured power is over 30 dB below the TE_{11} power. This is consistent with a selectivity value of greater than 30 dB for the TE_{12} coupler with respect to the TE_{11} mode, as was quoted earlier. For the 7-ripple case, the finite selectivity of the TE_{11} coupler comes into play, and gives an artificially high value for the measured TE_{11} power.

VI. Conclusions

A brief description of the method used to design mode-selective couplers has been presented, along with parameters for a TE_{11} and a TE_{12} coupler that have been fabricated and tested. Good agreement between predicted and actual performance has been obtained. These couplers were then used to determine the characteristics of other multimode devices, once again with good agreement between theory and experiment. These latter experiments also demonstrated the limitations that these couplers have due to the finite rejection of other modes. Couplers may be designed to detect other TE modes by a change in coupling guide and hole dimensions, while couplers for the TM modes would require coupling through the broad wall of the rectangular waveguide instead of the narrow wall.

Acknowledgments

References

1. Bhanji, A., Hoppe, D., Hartop, R., Stone, E. W., Imbriale, W. A., Stone, D., and Caplan, M., "High Power Ka-Band Transmitter for Planetary Radar and Spacecraft Uplink," *TDA Progress Report 42-78*, pp. 24-48, Jet Propulsion Laboratory, Pasadena, Calif., April-June 1984.

2. Miller, S. E., "Coupled Wave Theory and Waveguide Applications," *Bell System Technical Journal, 33*, pp. 661-719, 1954.

3. Hoppe, D., "An Experimental TE_{12}-TE_{11} Circular Waveguide Mode Converter," *TDA Progress Report 42-82*, pp. 43-56, Jet Propulsion Laboratory, Pasadena, Calif., April-June 1985.
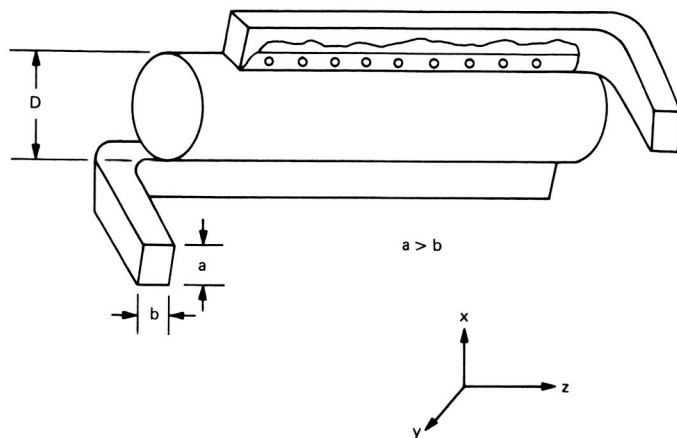
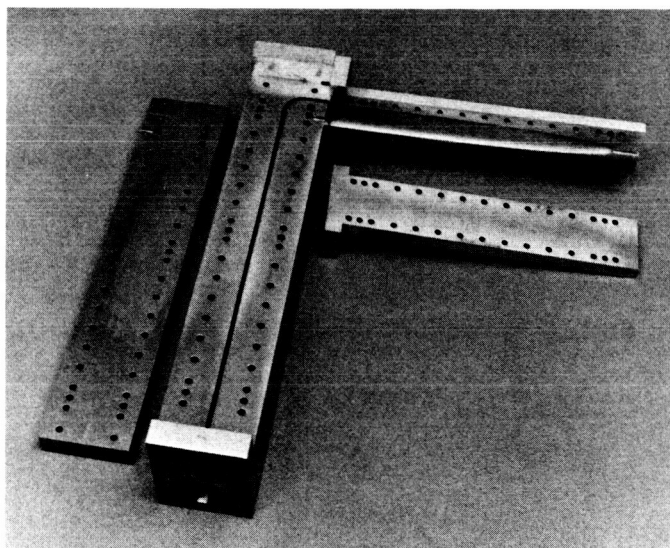Fig. 1. Mode-selective coupler configuration
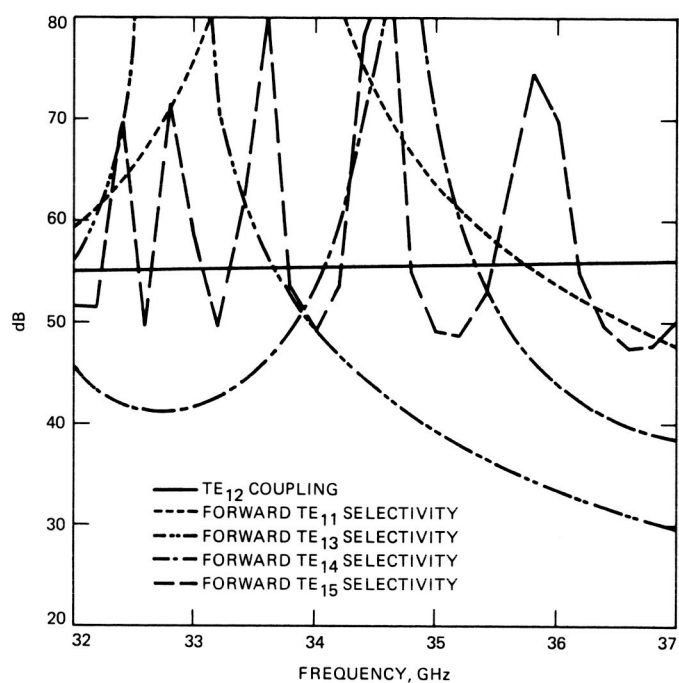

Fig. 2. Fabricated TE_{11} coupler

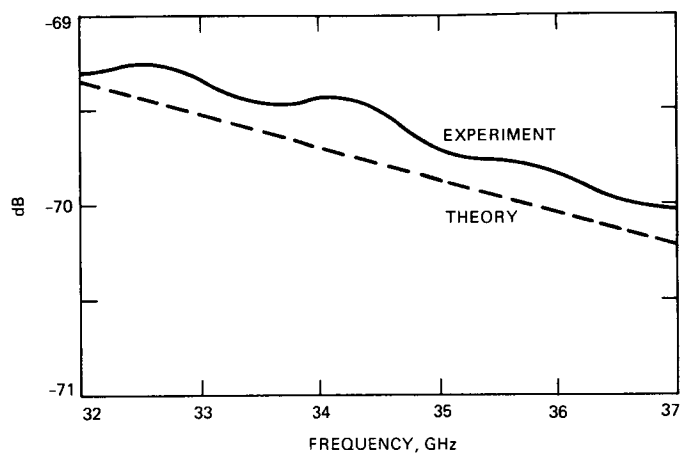Fig. 3. Predicted coupling and selectivities for TE_{12} coupler

15

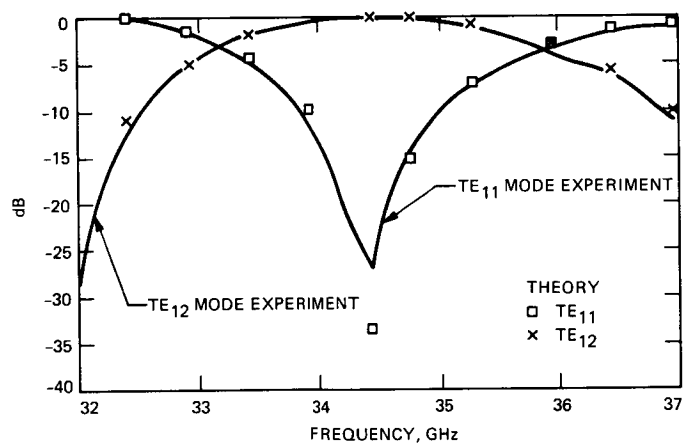Fig. 4. The TE_{11} coupling measurement vs theoretical results

Fig. 6. Mode converter mode content vs distance

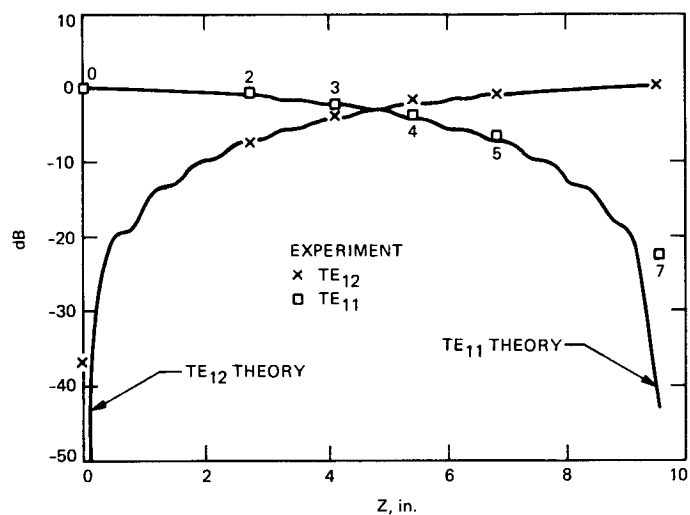Fig. 5. The TE_{11} to TE_{12} mode converter frequency response

Phase Noise Measurements of the 400-kW, 2.115-GHz (S-Band) Transmitter

P. Boss, D. Hoppe, and A. Bhanji

Radio Frequency and Microwave Subsystems Section

This article describes the measurement theory and presents a test method to perform phase noise verification using off-the-shelf components and instruments. The measurement technique described consists of a double-balanced mixer used as phase detector, followed by a low noise amplifier. An FFT spectrum analyzer is then used to view the modulation components. A simple calibration procedure is outlined that ensures accurate measurements. A block diagram of the configuration is presented as well as actual phase noise data from the 400-kW, 2.115-GHz (S-band) klystron transmitter.

I. Introduction

Various papers (Refs. 1 and 2) contain outlines of methods to measure transmitter phase stability. This article describes the implementation of the phase detector method, with particular emphasis on close-in phase noise measurement. Practical problems encountered during such a measurement are also described. Simple, accurate, and easily reproducible measurements are necessary so that improved transmitter designs can evolve.

Unwanted modulation and noise in the transmitted RF signal are caused by a variety of sources — specifically, shot noise in the klystron, fluctuations in the cathode beam power, RF drive power, and heater supply voltages, cooling water temperature fluctuations, focus supply fluctuations, and mechanically induced vibrations from cooling fans and water pumps. These pertubations will in most cases produce both angle and amplitude modulation. However, in klystron amplifiers, near the carrier, angle modulation (phase noise) predominates.

The most significant source of phase error is generally the cathode voltage applied to the klystron. As the cathode voltage increases and decreases, the corresponding increase and decrease in electron beam velocity create a decrease and increase in the apparent electrical length of the tube. This results in a higher quantity of electrons being emitted, which in turn produces an increase in amplitude of the RF output. Generally, as stated earlier, any amplitude disturbance is an order of magnitude less significant than the phase change and can therefore be ignored.

Another factor that introduces apparent change in phase length of the linear beam tube is the focusing magnet. As the electron beam enters the magnetic field, any component of the magnetic field which is perpendicular to electron velocity creates a transverse velocity component of the electrons. Since the energy of the electrons is conserved, any transverse velocity component that is added will detract from the axial velocity component, thus causing an apparent increase in the electrical length of the tube.

The RF drive amplitude variation is another factor which induces phase delay in the klystron amplifier; it is dependent on the tube gain (number of cavities), how close the tube is run to saturation, and the operating frequency of the tube. Since variation in drive power results in extracting more or less energy from the electron beam, there are corresponding changes in the beam velocity that induce phase changes.

Variations in heater voltages, which will change cathode temperature, will also have an effect on the phase error. Similarly, variations in the klystron body temperature will affect the length of the circuit and thus will also affect the phase error.

Finally, one must consider such effects as vibration in the tube from mechanical cavity tuner mechanisms that can physically move due to vibration or antenna movement. In addition, physical vibration of coaxial cables, couplers, and similar components should be considered.

II. Noise Spectrum

A frequency domain representation of a CW carrier with noise sidebands is shown in Fig. 1. The vertical scale represents the ratio of single sideband (SSB) power in a 1-Hz bandwidth, at a frequency fm away from the carrier, to the total signal power. This quantity is given the symbol $\mathcal{L}(f)$, with units dBc/Hz (dB below carrier in a 1-Hz bandwidth). Noise power varies directly with measured bandwidth. Thus, noise measured in a bandwidth other than 1 Hz may be normalized to a 1-Hz bandwidth by subtracting the correction factor given in Eq. (1).

$$\text{Correction Factor} = 10\log_{10} \frac{\text{BW}}{1\ \text{Hz}} \qquad (1)$$

where BW is the measured bandwidth in Hz. The power of discrete deterministic sidebands does not scale with measurement bandwidth; it remains constant.

As stated earlier, the phase noise in a klystron dominates *am* noise at frequencies close to the carrier, and to measure this noise the phase detector method was used.

III. Phase Detector Method

The theory behind this measurement is fairly straightforward. The basic components are a double balanced mixer used as a phase detector, a low noise amplifier, and a low frequency spectrum analyzer. The relationship between the RMS variation in phase between the LO and RF ports of the mixer and

the ratio of the power in a single sideband to the power in the carrier (P_{SSB}) is given by:

$$P_{\text{SSB}} = \frac{\phi^2}{2}$$

where ϕ is the difference between the RF and LO ports. Also, the relationship between the IF voltage (V_{IF}) of the mixer and ϕ is given by:

$$V_{\text{IF}} = K \cos\left(\pi/2 + \phi\ \text{rms}\right)$$
$$V_{\text{IF}} \cong K\ \phi\ \text{rms} \qquad (2)$$

where K is the mixer sensitivity and

$$V_m = A V_{\text{IF}}$$

where A is the LNA gain and

$$P_{\text{SSB}} = \frac{V_m^2}{2(AK)^2}$$

This is valid only for phase deviations of less than 0.1 radians.

Thus, by measurement of V_m and knowledge of the characteristics of the circuit (A and K), the noise spectrum can be determined.

IV. Hardware

A complete block diagram of the measurement system is shown in Fig. 2. Several elements must be added to the simple circuit described earlier to allow an accurate measurement. Circulators to provide a matched impedance, attenuators to control the power level, and couplers to monitor the power level should be placed at both the LO and RF ports of the mixer. A phase shifter to control the offset phase should be placed in the RF side of the mixer circuit. A bias tee should be placed at the IF port of the mixer to provide a non-reactive termination. It was also found necessary to place a 1 kohm resistor from the output of the bias tee to ground to increase the low frequency sensitivity of the mixer. In addition, a DC blocking capacitor was needed to prevent any DC offset from the mixer from forcing the high gain LNA to saturate. One should also consider the noise contribution of the test system elements. When selecting elements such as the driving source (i.e., the LO signal) and the low noise amplifier, care should be taken to choose generators and amplifiers that will not contribute significant noise compared to what is being measured. An HP 3561 A dynamic signal analyzer was used as the low frequency spectrum analyzer.

V. Calibration

A simple method of calibration was chosen so that it was not necessary to determine all the parameters of the measurement system. The system was calibrated by injecting a known phase-modulated signal and measuring the results. This is performed for all modulation frequencies of interest in order to form a calibration curve for the circuit at the frequencies of interest. For accurate results, it is important that the signal modulation be fairly pure phase modulation of a known P_{SSB} level. In calibrating the circuit, a P_{SSB} ratio of –30 dB was used with higher order sidebands at –50 dBc or lower. This level is set with an RF spectrum analyzer and modulating at a frequency that can be seen. The modulation frequency was then stepped from 100 kHz down to 0.01 Hz and the sideband observed with the low frequency FFT analyzer. For system calibration, the analyzer was used in the narrow-band mode to read the actual values of the modulation sidebands.

It was found that the major contributor to any frequency roll-off is the high-pass filter formed by the DC blocking capacitor and the combination of the corresponding circuit element resistances. This can be fairly well modeled by a simple RC filter whose corner frequency is about 2 Hz. The response also falls off a little at frequencies higher than 60 Hz, but since the main frequencies of interest for this measurement are below 20 kHz, this effect can be neglected. This response will then be used to correct the measured data.

VI. Measurement

When setting up the system to make a measurement, care must be taken to match the calibration conditions as closely as possible. The power levels seen by the mixer at its RF and LO ports are especially critical. Values of P_{LO} = +8 dBm and P_{RF} = –2 dBm were chosen for this measurement. These values can be changed but they must be repeated between calibration and an actual measurement.

The mixer must also be in quadrature for maximum sensitivity to phase variations and minimum sensitivity to amplitude variations. This is determined by monitoring the mixer IF voltage before the DC blocking capacitor. Because it can take hours to make a measurement at low frequency spans, it is desirable to know about any long-term phase drifts. An attempt was made to continuously monitor this point to check for very long-term phase drifts, but this led to problems because the spectrum analyzer was able to detect noise from the instrument used to monitor this point. Quadrature was then checked before and after the measurement in order to detect unacceptable amounts of drift.

The measurement was made at seven different frequency spans. The data for each of these spans was then combined to form one plot. These frequency ranges were:

0–100 kHz

0–10 kHz

0–1   kHz

0–100 Hz

0–10 Hz

0–1   Hz

0–0.1 Hz

The measurement data for each frequency span was read from the spectrum analyzer into the computer, where any gain or attenuation effects imposed by the system and determined in the calibration were weighted out. This provided results in the form of dBc/Hz. Each frequency span consists of 400 data points, which provides single-point resolution of 0.25 mHz on the 0.1 Hz frequency span.

When making this measurement, the spectrum analyzer will automatically range to accommodate the largest noise signal seen at its input. If there are relatively high-level discrete lines, the analyzer will choose a range to accommodate these signals. Since the dynamic range of the analyzer is 80 dB, low-level random noise cannot be detected when these discrete lines are present. Although the noise per Hz typically rises as one approaches the carrier, this effect is overwhelmed by the low frequency roll-off of the measurement system, and by the decreasing measurement bandwidth.

To avoid the above problem, which was caused by cathode voltage power supply lines and other discrete lines, a low-pass filter was used for spans less than 10 Hz. The filter had a corner frequency of about 2 Hz, and its effect was corrected for in the data measured at 1 Hz or less.

Since the spectrum analyzer was capable of storing the time data as well as the frequency data, the measurement was done using this "time buffer" capability. The time data can then be searched for any one-time event such as a step in phase. Events of this type can cause very deceptive results in the measured noise level. This method, although helpful, requires the storing and handling of large amounts of data.

VII. Measured Data

Figure 3 shows the results of the phase noise measurement made on the class C driver amplifier. It should be noted that this follows a theoretical $1/f$ noise characteristic. It is also

apparent that the driver noise is approximately 20 dB lower than the noise measured with the klystron in the circuit. This plot is shown in Fig. 4. The plot of the transmitter phase noise reveals several discrete spurious lines. Most of these components are traceable to the cathode voltage noise characteristics, which are shown in Fig. 5. The dominant lines are at 400 Hz due to the generator frequency, at 800 Hz (the 2nd harmonic), and at 2400 Hz (a product of the 400-Hz rectification). Also present is an unexpected line at 16 Hz, which was found to be a product of the cathode voltage supply's control feedback loop.

When the cathode supply voltage noise spectrum and the phase noise spectrum are compared, it becomes apparent that the two are very similar. This implies that the supply voltage is the major contributor to the noise. This is expected because the cathode voltage has the highest phase pushing factor of the affecting parameters. It is also seen that the noise of the system approaches the noise of the driver amplifier at high frequencies.

VIII. Conclusions

This article has presented a technique for measuring the phase noise of RF power amplifiers that is easily reproducible using off-the-shelf instruments and common electronic components. Measurement results are readily interpreted and unambiguous. This method will assist in the design evolution process (improved power supply regulation and long-term stability, improved temperature control of the cooling water, etc.) by correlating design changes with overall performance measurements.

Acknowledgments

References

1. Nocella, P. (Sperry Corporation), "Practical Purity Measurement of Pulse Doppler Radar Transmitter Amplifiers," *IEEE 17th Pulse Power Modulator Symposium, 1986*, pp. 32–38.

2. "Spectrum Analysis Noise Measurements," *Hewlett-Packard Application Note 150-4*, Palo Alto, California, April 1974.
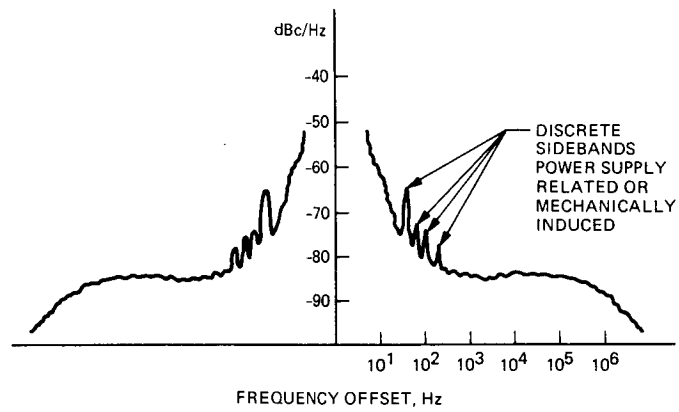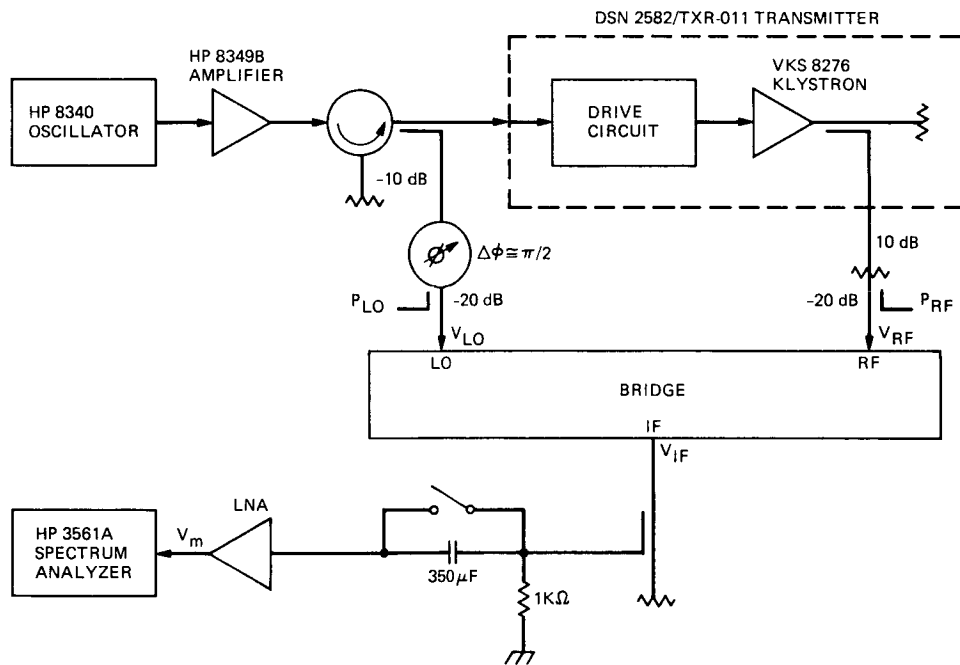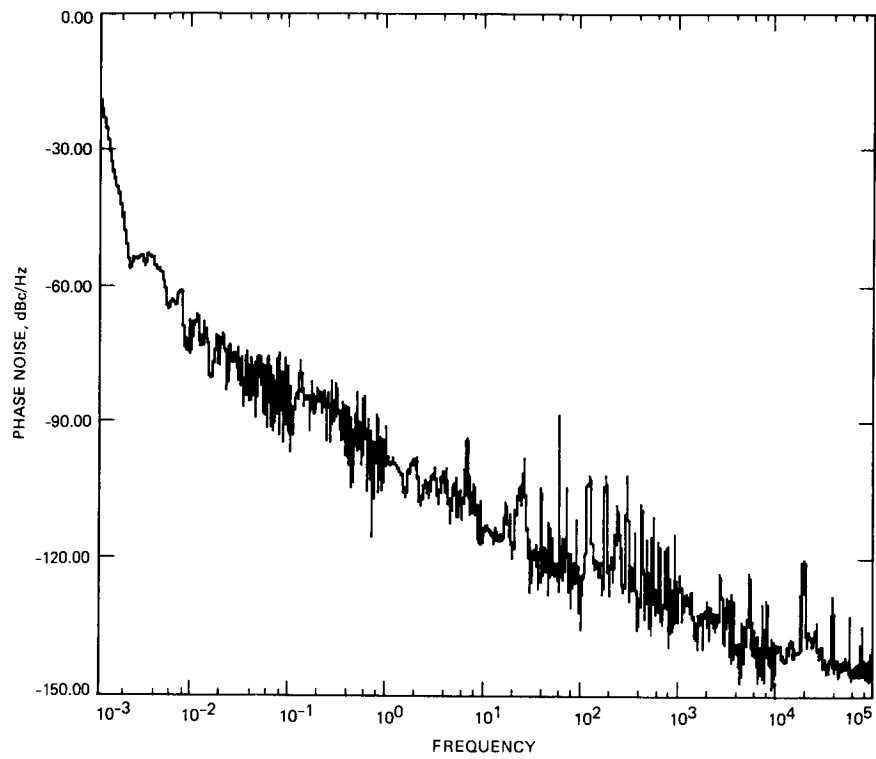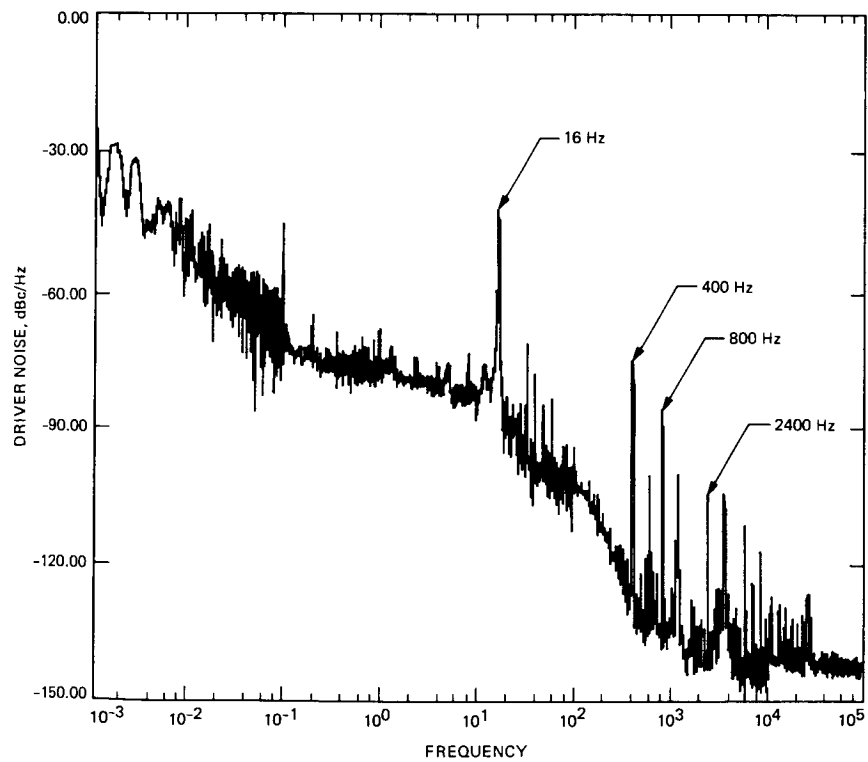
Fig. 1. The CW phase noise

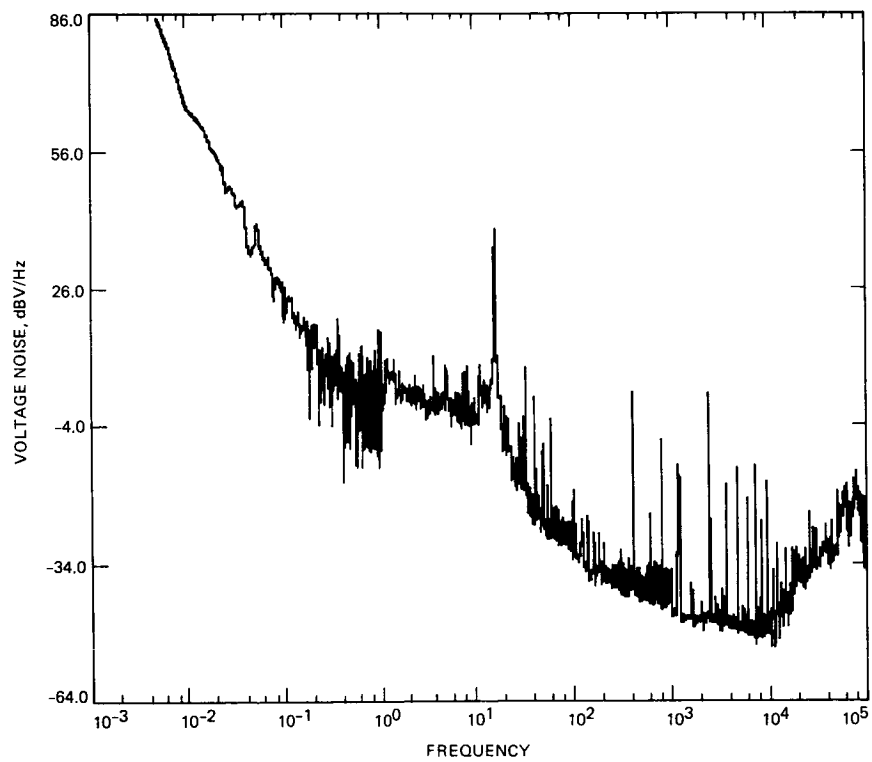**Fig. 2. Phase noise measurement system**

Fig. 3. Class C driver amplifier phase noise

Fig. 4. The 2.115-GHz, 400-kW transmitter phase noise

Fig. 5. Cathode voltage supply noise spectrum

Microwave Surface Resistance of Reactively Sputtered NbN Thin Films

J. J. Bautista

Radio Frequency and Microwave Subsystems Section

D. M. Strayer

Microgravity and Flight Experiments Research Section

The surface resistance of niobium nitride (NbN) thin films has been measured at 7.78 and 10.14 GHz in the temperature range of 1.5 to 4.2 K. The films were reactively sputtered on sapphire substrates to a thickness of approximately 1 micron. The surface resistance was determined by measuring the quality factor (Q) of the TE_{011} mode of a lead-plated copper cavity where the NbN served as one end-cap of the cavity.

I. Introduction

Although the B1 and A15 superconductors are of primary interest for the development of active devices (Josephson junction and SIS devices), homogenous, high critical temperature (T_c) superconductors are also of interest for passive microwave device applications. Dissipative losses and interfering signals can significantly increase the noise temperature of cryogenic ultra-low noise microwave amplifiers such as masers (Ref. 1) and High Electron Mobility Transistor (HEMT) amplifiers (Ref. 2). Microstrip and stripline superconducting filters and transmission lines can provide protection without significantly increasing the amplifier noise temperature (Ref. 3). Furthermore, the widespread availability of relatively inexpensive commercial, close-cycle refrigerators (10- to 15-K operating range) will readily allow the integration of high T_c superconducting devices with HEMTs.

For temperatures below 4.2 K, high T_c resonant structures can significantly improve the performance of maser amplifiers and cavity-stabilized oscillators. In a traveling wave maser (TWM) amplifier, dissipative losses in the slow-wave structure limit the amplifier noise temperature and gain. A superconducting slow-wave structure having a high second critical magnetic field would allow the construction of a TWM amplifier with near theoretical response, lowering the noise temperature by as much as 50% and increasing the gain bandwidth product by as much as 100%. The long-term frequency stability of a superconducting cavity-stabilized oscillator (SCSO) relies on the high Q and good mechanical stability of the superconducting cavity. A high T_c superconductor deposited on a sapphire substrate to form a superconducting cavity would allow operation of the oscillator at a higher temperature. For example, for a cavity having a Q of 10^9, 10^{-17} stability can be obtained using a lead film only by cooling to

0.9 K, while an NbN cavity would allow operation at 2 K (Ref. 4). Both these systems place requirements on the surface resistance of the superconducting film at microwave frequencies.

Surface resistance measurements can also serve as a deposition process monitor and as a non-destructive means of evaluating the homogeneity and quality of the film (Refs. 5 and 6). This work was initiated to determine the feasibility of the B1 compound NbN for passive microwave applications, since it can be fabricated with relative ease and, in addition, good quality thin films have been demonstrated using various deposition methods (Refs. 7, 8, and 9). This article presents the preliminary results of measurements of the surface resistance of NbN at X-band frequencies.

II. Sample Preparation

The NbN thin films were dc-magnetron sputtered at an approximate rate of 1700 A/min onto flat sapphire substrates to a thickness of about 1 micron. Throughout the deposition, the nitrogen partial pressure was maintained at 0.57 Pa and the argon at 1.56 Pa. The deposition was initiated with the substrates at room temperature and no attempt was made to control the substrate temperature during the deposition.

The epi-finished sapphire substrates were purchased from Insaco, Inc. (Quakertown, Penn.). Before the deposition, the substrates were cleaned with soapy de-ionized water and then rinsed in de-ionized water. The substrates were then degreased by rinsing them in acetone (5 min), methanol (5 min), and freon (5 min).

A variety of sample sizes were deposited for three different experiments. Three 5-cm diameter wafers (two 0.025- and one 0.300-cm thick) were made for the Q-cavity measurements presented in this report. A second set of four rectangular samples (7.00 by 2.03 by 0.064 cm) were made for calorimetric surface resistance measurements (Ref. 10). The third set of rectangular samples (four 9.40 by 0.56 by 0.076 cm and four 12.7 by 2.13 by 0.152 cm) were saved for slow-wave structure patterning and subsequent insertion loss measurements.

III. Microwave Measurements

The two cylindrical X-band cavities used for the Q measurements were made of OFHC copper electroplated with lead to a thickness of approximately 3 microns (Ref. 11). The cavities have diameter to length ratios of 1.0 and 1.75, and were operated in the TE_{011} mode at 7.78 and 10.14 GHz, respectively. The cavity Q's were all determined by pulsing the incident microwave energy and measuring the decay time of the stored cavity energy (Ref. 12). For these measurements, the homogeneous Pb(Cu) cavities were first characterized as a function of temperature and the surface resistance (R_s) was determined using the following expression:

$$R_s = \frac{G}{Q_o} \qquad (1)$$

where G is a geometrical factor (780 and 619 ohms, respectively, for the present geometries) and Q_o is the baseline quality factor for the Pb(Cu) surface only.

The surface resistance at 7.78 GHz of samples 1 and 2 (NbN on 0.025-cm-thick substrates) was determined by placing them on top of the bottom end-cap of the cavity and repeating the decay time measurements. At 10.14 GHz, sample 3 (NbN on 0.300-cm thick substrate) replaced the removable Pb(Cu) end-cap. For this cavity configuration, the relationship between the surface resistance of the thin-film end-cap (R_x) and the measured quality factor, Q_x, is given by the following expression:

$$R_x = \frac{\left[1 + \frac{\left(\frac{Q_o}{Q_x} - 1\right)}{f}\right] G}{Q_o} \qquad (2)$$

where f (0.072 and 0.206, respectively) is the fraction of the total cavity loss associated with the thin-film end-cap.

IV. Experimental Results and Discussion

The results at 7.78 GHz for samples 1 and 2 are plotted in Fig. 1. Included for comparison are the Stanford calorimetric data (Ref. 10) for sample 4 at 8.6 GHz. Although the Q-cavity and calorimetric cavity data appear to be in agreement at temperatures near 4 K, the measured results for samples 1 and 2 show lots of scatter and appear to be temperature independent. Since samples 1 and 2 rest on top of the cavity end-cap, it is strongly suspected that inadequate heat contact with the helium bath resulted in the observed thermal heating effects. To obviate this latter difficulty, the 10.14-GHz cavity was constructed so that it would accept sample 3 as the end-cap. This configuration results in a sample that is in direct contact with the helium bath. Figure 2 contains a plot of the results for NbN sample 3, along with the data for the Pb(Cu) at 10.14 GHz. Both sets of data clearly exhibit a strong temperature dependence; however, the expected (BCS) precipitous drop in surface resistance is absent.

The data for sample 3 was fit by linear regression to a linearized form of the BCS surface resistance for $T < 0.5T_c$ given by:

$$R_s(T) = A \exp\left[-e_g\left(\frac{T_c}{T}\right)\right] + R_o \qquad (3)$$

where R_o is a residual surface resistance, e_g is the normalized zero temperature energy gap parameter, and T_c is the critical transition temperature. Figure 3 contains a plot of the best fit to the data of sample 3, assuming $T_c = 15.2$ K and $R_o = 10$ micro-ohms. The resulting 0.52 value of e_g is a factor of one half smaller than the smallest values determined from tunneling measurements (Ref. 9).

Although the e_g value for NbN sample 3 is anomalously small, it is in agreement with the results obtained by Isagawa at 6.5 GHz on samples of NbN reactively sputtered on Nb (Ref. 13). But in contrast to the NbN results, the linear regression analysis for the 7.78- and 10.14-GHz Pb(Cu) cavity measurements yielded e_g values of 1.98 and 2.00, respectively, which are approximately 10% less than the tunneling measurement values (Ref. 14).

The large difference between the e_g values determined by RF measurements and those determined by dc tunneling measurements is believed to be associated with the columnar structure of sputtered NbN and the compounds that form between the NbN and the 10 A Nb_2O_5 outermost layer. The smaller e_g value obtained from RF measurements is the result of the formation of a metallic NbN suboxide with a T_c of approximately 7 K adjacent to the serrated NbN (Ref. 15). The dc tunneling currents flow directly across the serrated peaks of NbN, while RF currents are induced in both the NbN and the lossier NbN suboxide, resulting in an averaged lower e_g.

V. Conclusions

In Fig. 1, for comparison to a normal metal, the surface resistance is plotted for OFHC Cu, calculated from the insertion loss measurements as a function of temperature for an interdigital filter. Comparison of this Cu data with the calorimetric data shows that below 12 K, the NbN surface resistance is at least 1/10th the value of Cu. Future planned insertion loss measurements of NbN slow-wave structures at 2.3 and 8.4 GHz subjected to large dc magnetic fields will determine the applicability of NbN to TWMs. Although the question regarding the applicability of NbN for SCSOs remains unresolved, it is evident that this material is clearly applicable for microstrip and stripline devices operating in the 10- to 15-K temperature range.

References

1. Bautista, J. J., and Petty, S. M., "Cryogenic Filters for RFI Protection," *TDA Progress Report 42-65*, p. 94, July–August, 1981.

2. Pospieszalski, M. W., Weinreb, S., Chao, P. C., Mishra, U. K., Palmeter, S. C. Smith, P. M., and Hwang, J. C. M., "Noise Parameters and Light Sensitivity of Low-Noise HEMTs at 300 and 12.5 K," *IEEE Trans. Electron Devices, ED-33*, pp. 218–223, February 1986.

3. Bautista, J. J., and Petty, S. M., "Superconducting NbTi and Pb(Cu) Bandpass Filters," *IEEE Trans. Magn. MAG-21*, No. 2, pp. 640–643, 1985.

4. Strayer, D. M., Dick, G. J., and Mercereau, J. E., "Performance of a Superconducting Cavity Stabilized Ruby Maser Oscillator," *Applied Superconductivity Conference Paper No. MH-4, 1986 Applied Superconductivity Conference*, September 29–October 3, 1986.

5. Allen, L. H., Beasley, M. R., Hammond, R. H., and Turneaure, J. P., "RF Surface Resistance of High T_c Superconducting A15 Thin Films," *IEEE Trans. Magn.*, *MAG-21*, No. 3, p. 1003, 1983.

6. Allen, L. H., Anklam, W. J., Beasley, M. R., Hammond, R. H., and Turneaure, J. P., "RF Surface Resistance in Nb_3Sn Thin Films," *IEEE Trans. Magn., MAG-21*, No. 2, pp. 525–527, 1985.

7. Villegier, J. C., Vieux-Rochaz, L., Goniche, M., Renard, P., and Vabre, M., "NbN Tunnel Junctions," *IEEE Trans. Magn., MAG-21*, No. 2, pp. 498–504, 1985.

8. Gurvitch, G., Remekia, J. P., Rowell, J. M., Geerk, J., and Lowe, W. P., "Tunneling, Resistive and Structural Study of NbN and Other Superconducting Nitrides," *IEEE Trans. Magn., MAG-21*, No. 2, pp. 509–513, 1985.

9. Bacon, D. D., English, A. T., Nakahara, S., Peters, F. G., Schreiber, H., Sinchair, W. R., and van Dover, R. B., "Properties of NbN Thin Films Deposited on Ambient Temperature Substrates," *J. Applied Phys., 54*, pp. 6509–6516, 1983.

10. Allen, L. H., Beasley, M. R., and Turneaure, J. P., "RF Surface Resistance of Nb_3Sn, NbZr, and NbN Thin Films," *ASC Paper No. MG-20, 1986 Applied Superconductivity Conference*, September 29–October 3, 1986.

11. Dick, G. J., and Delayen, J. R., "A New Chemical Polishing Procedure for Lead-Plated Copper Superconducting Accelerating Resonators," *IEEE Trans. Magn., MAG-19*, No. 3, 1983.

12. Yogi, T., *Radio Frequency Studies of Surface Resistance and Critical Magnetic Field of Type I and Type II Superconductors*, Ph.D. Thesis, California Institute of Technology, 1977.

13. Isagawa, S., "RF Superconducting Properties of Reactively Sputtered NbN," *J. Applied Phys., 52*, pp. 921–927, 1981.

14. Kittel, C., *Introduction to Solid State Physics*, 5th Ed., John Wiley and Sons, NY, NY, pp. 412–413, 1976.

15. Halbritter, J., "Superconductivity of Oxidized NbN Compound Surfaces," *ASC Paper No. MC-11, 1986 Applied Superconductivity Conference*, September 29–October 3, 1986.
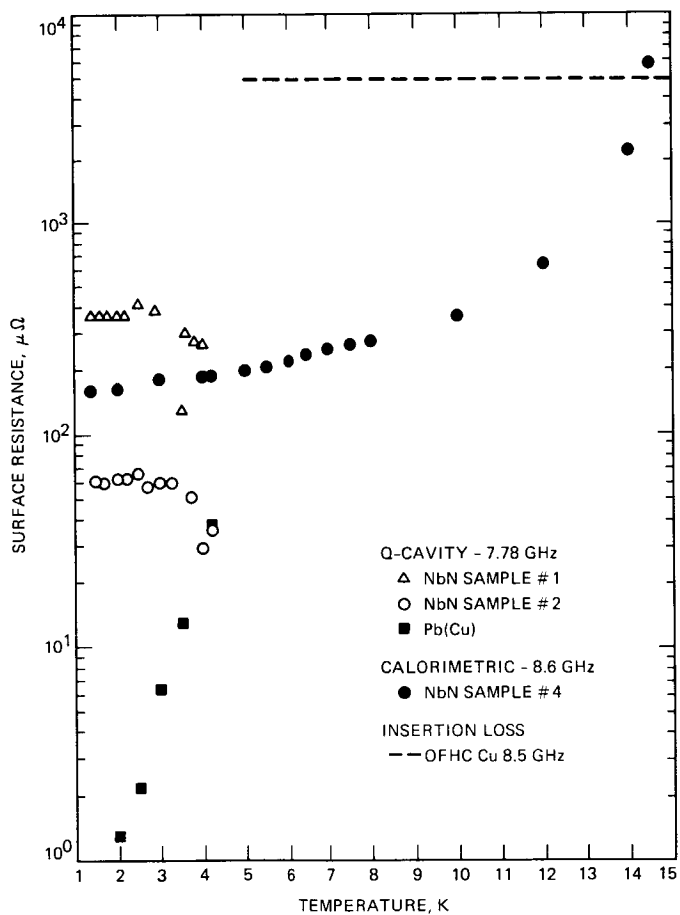
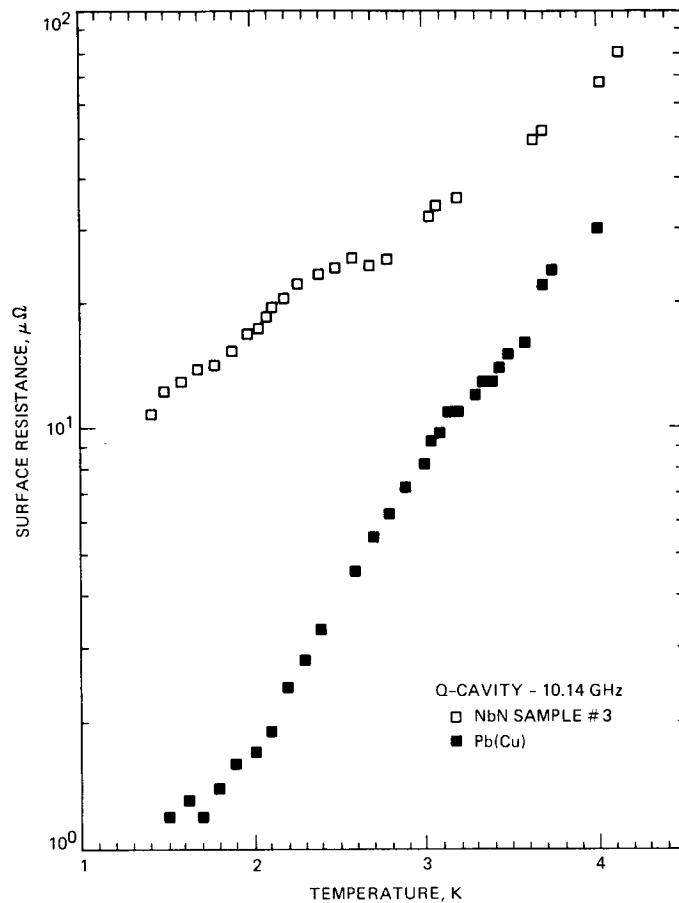**Fig. 1.** Surface resistance measurements of NbN samples at 7.78 and 8.6 GHz

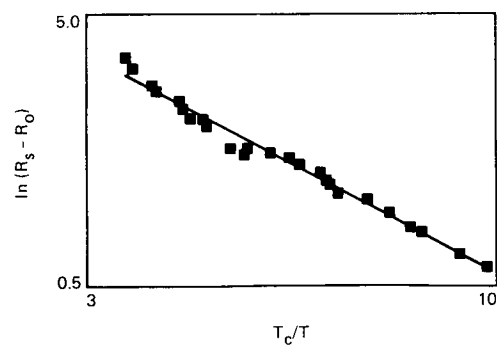**Fig. 2.** NbN surface resistance measurements at 10.14 GHz

Fig. 3. Linear regression fit to sample 3

Statistical Analysis of the 70 Meter Antenna Surface Distortions

K. Kiedron, C. T. Chian, and K. L. Chuang
Ground Antennas and Facilities Engineering Section

Statistical analysis of surface distortions of the 70 meter NASA/JPL antenna, located at Goldstone, was performed. The purpose of this analysis is to verify whether deviations due to gravity loading can be treated as quasi-random variables with normal distribution. Histograms of the RF pathlength error distribution for several antenna elevation positions have been generated. The results indicated that deviations from the ideal antenna surface are not normally distributed. The observed density distribution for all antenna elevation angles is taller and narrower than the normal density, which results in large positive values of kurtosis and a significant amount of skewness. The skewness of the distribution changes from positive to negative as the antenna elevation changes from zenith to horizon.

I. Introduction

The surface of the main reflector of a Cassegrain antenna is supposed to be a perfect paraboloid. In practice, however, there are systematic and random causes of surface irregularities. Ripples on the antenna surface as shown in Fig. 1 are basically a product of all gravity, wind and thermal loading effects. Additionally, some distortions are introduced at the factory during panel manufacturing, and some are caused by imperfect panel alignment and setting.

The gain loss, ΔG, due to reflector surface imperfections can be computed in decibels from:

$$\Delta G = 10 \log_{10} \eta \qquad (1)$$

where the efficiency η is approximated by Ruze's (Ref. 1) equation as:

$$\eta = \exp -\left(\frac{4\pi\, rms}{\lambda}\right)^2 \qquad (2)$$

where the root mean square (*rms*) of surface deviations is defined as one-half the change in the RF pathlength and λ is the operating wavelength (Refs. 2, 3). Fig. 2 illustrates the geometrical relationship, between an arbitrary surface displacement d, normal deviation n, and the corresponding RF pathlength deviation δ.

Ruze's Eq. (2) is derived under the assumption that the surface deviations δ_i at any surface point i are random variables having a Gaussian distribution with a zero mean and a standard deviation equal to the *rms* of the surface deviations of the reflector $(\delta_i \epsilon N(0, rms))$. Eqs. (1) and (2) are commonly used to estimate antenna gain loss even when surface distortions are due to deterministic causes such as gravity.

The purpose of this article is to verify whether deviations due to gravity can be treated as quasi-random variable with normal distribution. In order to eliminate truly random deviations (due to panel misalignment, wind turbulence and local thermal effects as well as measurement uncertainty) surface deviations simulated by a mathematical finite element model of antenna structure are used.

II. Data Description

The pathlength errors δ_i, due to gravity loading, measured from the least square best fit paraboloid were obtained for the 70 meter antenna from a finite element model generated with the JPL-IDEAS structural program. Five different antenna elevation positions, 0, 30, 60, 75 and 90 degrees, have been analyzed. Rigging position, that is 45 degrees, is excluded from analysis, because rms is zero at that elevation. Due to antenna symmetry, only one half of the dish, represented by $N = 715$ points, was modeled. The rms (root mean square) is calculated as:

$$rms = \sqrt{\sum_{i=1}^{N} \delta_i^2 \, \omega_i} \qquad (3)$$

where N is the number of surface points and surface weighting factors w_i are proportional to the point tributary areas. The surface weighting factors w_i are normalized to meet the condition:

$$\sum_{i=1}^{N} w_i = 1 \qquad (4)$$

Statistical analysis is performed on a new rescaled variable:

$$X_i = \delta_i \, \alpha_i \qquad (5)$$

where

$$\alpha_i = \sqrt{w_i N}$$

Such a scaling allows direct comparison of the standard deviation of X and rms. The mean of X is approximately equal to zero (in all cases it was less than 0.013 mm) and the standard deviation agrees with the rms with better than 1% accuracy.

III. Test of Normality

For each antenna elevation position, a histogram of relative frequency distribution has been generated ($K = 30$ classes in -1.5 to 1.5 millimeter interval). Relative frequencies (proba-

bilities in $\Delta x = 0.1$ millimeter segments) are obtained by division of each class frequency by the total number of points $N = 715$. The mean, m, and standard deviation, σ, for the particular elevation position have been calculated. Histograms with the normal counterparts $N(m, \sigma)$ are shown on Figs. 3, 4, 5, 6 and 7 for the elevation angles 0, 30, 60, 75, and 90 degrees respectively. In all five cases, histograms do not follow the normal distribution. The departure from normality is verified by computing the value of chi square χ^2 statistics according to the expression:

$$\chi^2 = N \sum_{i=1}^{K} \frac{(p_i - P_i)^2}{P_i} \qquad (6)$$

where K is the number of classes, p_i is the probability of occurence of observed data in the i-th interval $(x_i, x_i + \Delta x)$ and P_i is the probability obtained from a hypothetical normal distribution $N(m, \sigma)$:

$$P_i = \frac{1}{\sqrt{2\pi} \sigma} \int_{x_i}^{x_i + \Delta x} \exp[-(x - m)^2 / 2\sigma^2] \, dx \qquad (7)$$

The hypothesis that the observed distribution is normal can be rejected with probability higher than 0.99 due to the large values of χ^2/N (see Table 1). The same results can be obtained by using GFIT program from IMSL (International Mathematical and Statistical Library) which computes the goodness of fit by χ^2 statistics, but it uses equiprobable categories (Ref. 4).

Since it is possible that the estimate of m and σ are incorrect it was checked whether values of χ^2 statistics can be decreased for other normal distributions with different m and σ. Actually χ^2 was minimized with respect to m and σ. This resulted in finding a normal distribution which best fits the observed data. The results, after the best fit, are presented in Table 2. Again it can be concluded that with probability higher than 0.98, hypothesis of normality, must be rejected.

IV. Non-Normality Characterization

Since it was verified that gravity induced path length error is not normal for any antenna elevation position it is desirable to characterize the observed type of non-normality (Ref. 5). The two, skewness and kurtosis, parameters characterize the type of non-normality. Using higher moments about the mean: (with $k > 1$)

$$m_k = \frac{1}{N} \sum_{i=1}^{N} (X_i - m)^k, \qquad k = 2, 3, 4 \qquad (8)$$

The coefficient of skewness is estimated according to the formula:

$$g_1 = \frac{m_3}{(m_2)^{3/2}} \tag{9}$$

where m_2 and m_3 are the second and third moments, respectively. If the sample comes from the normal population g_1 is approximately normally distributed with mean zero and standard deviation equal $\sqrt{6/N}$. The skewness measures the amount of symmetry around the mean. If low values of X are bunched close to the mean but high values extend far above the mean, this measure will be positive, since the large positive contributions of $(X - m)^3$, when X exceeds m, will predominate over the smaller negative contributions of $(X - m)^3$ obtained when X is less than m. In the population with negative skewness, the lower tail is the extended one, shifted to the left. In the population with positive skewness, the upper tail is the extended one, shifted to the right. The results of skewness for all elevation positions are shown in Table 3. The largest negative skewness is confirmed for the antenna at the horizon. Skewness decreases when the antenna approaches the rigging position (45 degrees). The largest positive skewness occurs for 60 degrees elevation and it decreases when the antenna approaches zenith position. The same trend can be observed on the histograms of Figs. 3, 4, 5, 6, and 7.

The degree of flattening of a probability density function near its mean is measured by the coefficient of kurtosis computed according to:

$$g_2 = \left(\frac{m_4}{m_2^2}\right) - 3 \tag{10}$$

In large samples, g_2 is normally distributed with mean zero and standard deviation equal $\sqrt{24/N}$. Positive values indicate that the probability density function is more sharply peaked, that is taller and narrower, than the normal density. Flat-topped distributions show negative kurtosis. The kurtosis parameter for all antenna elevation positions is presented in Table 3. Since the sample's values of g_2 are much larger than the standard error $\sqrt{24/N}$, the positive kurtosis parameter is confirmed for every antenna elevation position. Again, this result of exceeding the normal peak is noticeable on the histograms of Figs. 3, 4, 5, 6, and 7.

V. Summary

The statistical analysis of the 70 meter antenna surface distortions indicated that deviations from ideal paraboloid are not normally distributed. The distribution for all elevation angles is taller and narrower than normal density, which results in large positive values of the kurtosis and a significant amount of skewness. The skewness of the distribution changes from negative to positive as the antenna elevation angle changes from horizon to zenith. Therefore, the gravity induced deformations violate the assumption of normality in Ruze's equation. It is reasonable, however, to assume that for small departures from normality Ruze's equation will provide a good estimate of antenna gain loss. Nevertheless, the following open problem is posed: what modifications, if any, should be applied to Ruze's equation to incorporate non-zero skewness and kurtosis as a function of antenna elevation positions? Therefore Ruze's formula should be verified for non-normal quasi-random distribution by comparing it with results produced by the exact expression for antenna efficiency in the theory of electromagnetic scattering and diffraction (Ref. 6).

Acknowledgment

References

1. Ruze, J., "The Effect of Aperture Errors on the Antenna Radiation Pattern," *Nuevo Cimento Suppl.*, Vol. 9, No. 3, pp. 364–380, 1952.

2. Utku, S., and Barondess, S. M., "Computation of Weighted Root-Mean-Square of Path Length Changes by the Deformations and Imperfections of Rotational Parabolidal Antennas," *JPL Technical Memorandum 33-118*, Jet Propulsion Laboratory, Pasadena, Calif., March 1963.

3. Kiedron, K., "Homology Deviation Measure for Microwave Antenna Design," *IEEE Montech '86 Conference on Antennas and Communications*, Montreal Canada, pp. 173–176, September 1986.

4. Kendall, M. G., and Stuart, A., "*The Advanced Theory of Statistics*," Vol. 2, Hafner Publishing Company, New York, 1961.

5. Snedecor, G. W., and Cochran, W. G., "*Statistical Methods*" Sixth Edition, The Iowa State University Press, Ames, Iowa, 1967.

6. Rahmat-Samii, Y., "An Efficient Computational Method for Characterizing the Effects of Random Surface Errors on the Average Power Pattern of Reflectors," *IEEE Transactions Antennas and Propagation*, *AP-31*, No. 1, January 1983.

Table 1. Chi square goodness of fit test: m and σ one obtained with standard estimators

Elevation angle	m	σ	χ^2/N
0	−0.013	0.646	0.188
30	−0.003	0.204	0.600
60	0.002	0.190	0.239
75	0.002	0.367	0.183
90	0.001	0.534	0.090

Table 2. Chi square goodness of fit test: m and σ minimize x^2

Elevation angle	m	σ	χ^2/N
0	0.059	0.576	0.145
30	−0.013	0.239	0.233
60	0.008	0.215	0.156
75	−0.010	0.407	0.134
90	−0.003	0.474	0.069

Table 3. Skewness and kurtosis test

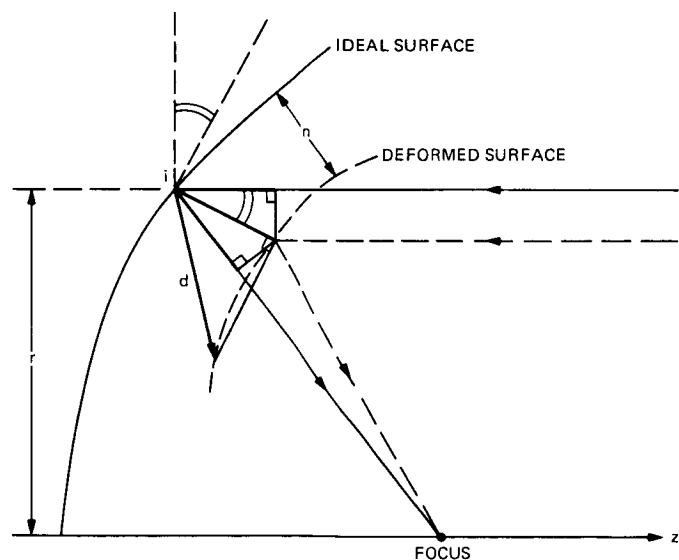| Elevation angle	Skewness	Kurtosis
0	−1.338	2.77
30	−0.972	1.82
60	0.568	1.61
75	0.429	2.16
90	0.363	2.99
$\sqrt{6/N}$	0.092	
$\sqrt{24/N}$		0.183
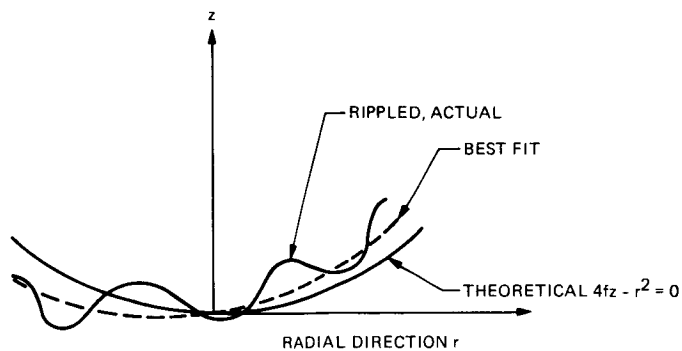
Fig. 1. Surface of main reflector

Fig. 2. Geometrical relationship between surface displacement d, normal deviation n, and corresponding RF pathlength deviation σ
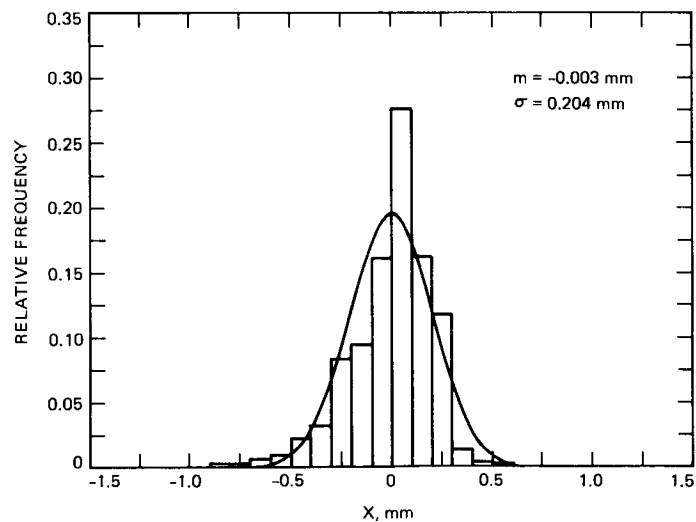

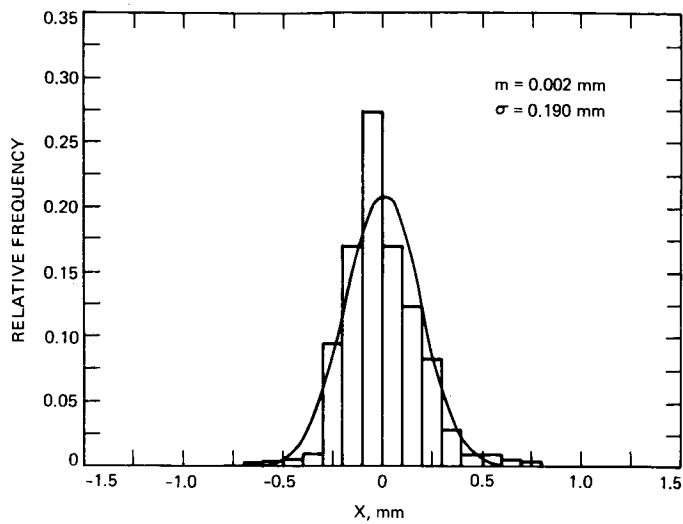Fig. 3. Histogram of RF pathlength error distribution at 0 degree elevation
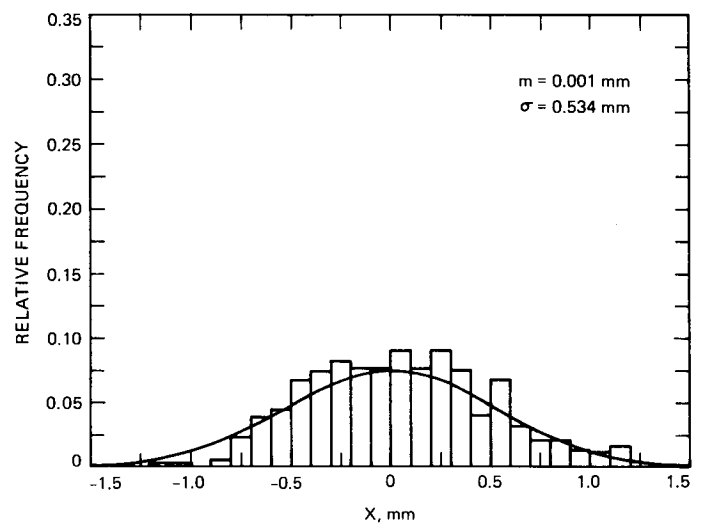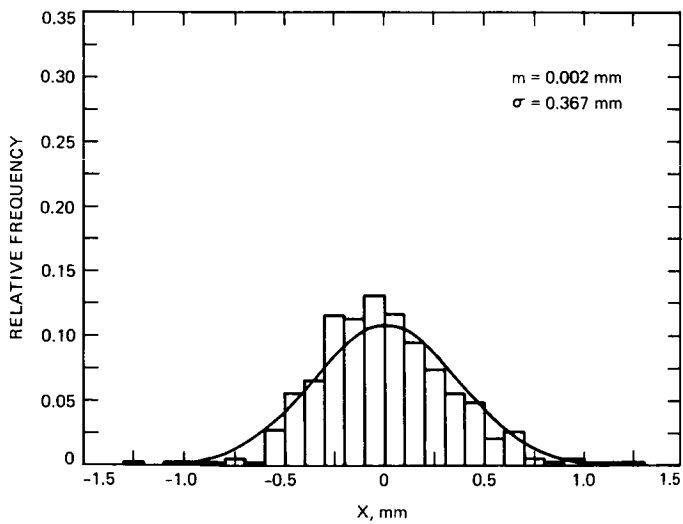

Fig. 4. Histogram of RF pathlength error distribution at 30 degree elevation

34

Fig. 5. Histogram of *RF* pathlength error distribution at 60 degree elevation

Fig. 7. Histogram of *RF* pathlength error distribution at 90 degree elevation

Fig. 6. Histogram of *RF* pathlength error distribution at 75 degree elevation

Antenna Pointing Systematic Error Model Derivations

C. N. Guiar and F. L. Lansing
Ground Antenna and Facilities Engineering Section

R. Riggs
TDA Engineering Office

The pointing model used to represent and correct systematic errors for DSN antennas is presented. Analytical expressions are given in both azimuth-elevation (az-el) and hour angle-declination (ha-dec) mounts for RF axis collimation error, encoder offset, nonorthogonality of axes, axis plane tilt, and structural flexure due to gravity loading. While the residual pointing errors (rms) after correction appear to be within the ten percent of the half-power beamwidth criterion commonly set for good pointing accuracy, the DSN has embarked on an extensive pointing improvement and modeling program aiming toward an order of magnitude higher pointing precision.

I. Introduction

Since the beamwidth of the parabolic reflector antenna decreases with increasing aperture size and decreasing wavelengths, errors in pointing accuracy assume additional importance. The potential increase of gain is rapidly lost if the narrower beam is not directed at the target with acceptable accuracy. The antennas being used by NASA-JPL for deep space tracking at Goldstone (California), Spain, and Australia operate at the S (2.3 GHz) and X (8.4 GHz) frequency bands. For a 70-m antenna operating at these frequencies, this translates into half-power beamwidths of 128 and 35 millidegrees, respectively. In order to effectively utilize these antennas at these sizes and frequencies, one should be able to point the RF axis to within 10 percent of the half-power antenna beamwidth.

A pointing model is commonly used to represent and correct systematic errors that contribute to pointing inaccuracies. Systematic errors are repeatable errors that are not determined by chance but by a bias. Random (stochastic) errors remain uncorrected and represent the "final" pointing system inaccuracy. This article describes the present mathematical model being used throughout the DSN and discusses possible future improvements for the model and the model-fitting process.

II. Sources of Systematic Pointing Error and System Overview

The antenna is commanded to point its beam axis at a (predicted) spacecraft location, referred to as a "predict" throughout the remainder of this article. The predicts are generated in the Antenna Pointing Assembly (APA), as shown in Fig. 1. Due to the many inherent pointing errors (systematic and random), the antenna boresight does not point to the desired point. Pointing corrections are therefore added to the predicted azimuth and elevation angles (or hour angle and declination angle) to compensate for the systematic portion of the pointing errors. These corrections are contained in the systematic error correction table of the Antenna Control Subassembly (ACS). Also located in the ACS are two other point-

ing corrections: (1) the refraction correction algorithm, which supplies an elevation correction based on temperature, humidity, and pressure values (Ref. 1) and (2) the "squint" correction table, which corrects for the boresight shift due to lateral motion of the subreflector. After the systematic, refraction, and "squint" corrections are added, the new command angles are sent to the Antenna Servo Controller (ASC) which generates the appropriate rate commands. During "blind" pointing, the shaft encoders send the feedback position signal for comparison with the command position signal. During spacecraft tracking and radio astronomy experiments, a different technique is used for closing in on the target whereby the antenna is scanned around its boresight in a circular pattern with a constant angular offset in order to detect the position of peak gain. This technique is referred to as conical scanning (conscan) (Ref. 2).

The key sources of systematic pointing error to date can be itemized as follows (Refs. 3, 4, and 5): (1) RF collimation error, (2) encoder fixed offset, (3) axis skew, (4) azimuth axis tilt, (5) structural flexure due to gravity, and (6) residual refraction error. Each error will be further defined for the two types of DSN antenna mounts (the azimuth-elevation mount and the hour angle-declination mount) in the paragraphs that follow. The total error in pointing is expressed as the algebraic sum of the individual error source contributors. Noting that the amount of error from most of the above error sources is a function of the angular position of the antenna, the total error in each coordinate is computed accordingly.

The relevant coordinates for an Az-El mounted antenna are azimuth and elevation, the errors being expressed in ΔAz and ΔEl. For a polar mount antenna, the errors may be expressed as ΔHA and ΔDec. Using these coordinates, it can be shown that a fixed pointing error quantified in terms of ΔAz for an Az-El mount (or ΔHA for a polar mount) will vary with the elevation (or declination) angle of the beam. This occurs since the beam is never in the azimuth (or hour angle) plane except at zero elevation (or declination) angles. In order to avoid this difficulty, a coordinate system is used where the beam itself forms one axis, with the other two axes forming a plane orthogonal to the beam. For an Az-El mount, one axis is the elevation axis, the other is normal to the beam and is contained by the elevation plane. This will be referred to as the cross-elevation axis (Fig. 2). Pointing errors in this coordinate system may be expressed as

Errors in the elevation plane = ΔEl

Errors in the cross-elevation plane = ΔXEl
$$= \Delta Az \times \cos (el)$$

Magnitude of total pointing error $\cong \sqrt{(\Delta El)^2 + (\Delta XEl)^2}$

and for a polar mount antenna as

Errors in the declination plane = ΔDec

Errors in the cross-declination plane = $\Delta XDec$
$$= \Delta Ha \times \cos (Dec)$$

Magnitude of total pointing error $\cong \sqrt{(\Delta Dec)^2 + (\Delta XDec)^2}$

III. Analysis of Systematic Errors

A. Azimuth (Hour Angle) Collimation Error

Collimation error is defined to be the nonorthogonality of the antenna beam to the elevation (or declination) axis. In order to derive an expression for collimation error, suppose a star is located at position X of the spherical system in Fig. 3 (Ref. 4) and the antenna is located in the center of the sphere. The antenna beam is displaced by an amount δ due to collimation error and points to position X'. The resulting azimuth error can be derived by solving spherical triangle WZX'

$$\cos (90 + \delta) = \cos (90 - el) \cos 90$$
$$+ \sin (90 - el) \sin 90 \cos (90 + \Delta Az) \tag{1}$$

which reduces to

$$-\sin \delta = \cos (el) \sin \Delta Az \tag{2}$$

If δ and ΔAz are small, Eq. (2) can be approximated by

$$\Delta Az = -\delta \sec (el) \tag{3}$$

When converted to XEl coordinates, the collimation error becomes

$$Xel = -\delta \tag{4}$$

There is no first order correction to elevation in the case of this azimuth collimation error.

Collimation error for a Ha/Dec-type mount is defined in the manner detailed above, with ha and dec replacing az and el, respectively.

B. Axis Skew

Axis skew is defined to be the nonorthogonality of the elevation (or declination) axis to the azimuth axis (or hour angle) axis. Figure 4 (Ref. 4) depicts the elevation axis skewness in the spherical system. The antenna is located in the center of the sphere with its azimuth and elevation axes

denoted by X_3 and X_2, respectively. The axes are skewed by an amount ξ, so the antenna points to the displaced position S' instead of the star position S. The relationship between the skewed and nonskewed axes is given by

$$[X'] = [B] [X] \tag{5}$$

where $[B]$ is the transformation matrix and is given as

$$[B] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\xi & -\sin\xi \\ 0 & \sin\xi & \cos\xi \end{bmatrix} \tag{6}$$

If ξ is small, $[B]$ becomes

$$[B] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\xi \\ 0 & \xi & 1 \end{bmatrix} \tag{7}$$

Equation (5) expands into the following relationship

$$\left.\begin{aligned} X_1' &= X_1 \\ X_2' &= X_2 - X_3\xi \\ X_3' &= X_2\xi + X_3 \end{aligned}\right\} \tag{8}$$

Axis skewness causes an error in both azimuth and elevation. The azimuth error due to skewness will be defined to be

$$\Delta Az = Az' - Az \tag{9}$$

For small values of azimuth error, Eq. (9) can be written as

$$\Delta Az = \tan Az' - \tan Az \tag{10}$$

where

$$\tan Az = \frac{X_2}{X_1} \tag{11}$$

and

$$\tan Az' = \frac{X_2'}{X_1} = \frac{X_2 - X_3\xi}{X_1} \tag{12}$$

The resulting ΔAz is

$$\Delta Az = -\frac{X_3}{X_1}\xi \tag{13}$$

Since X_3/X_1 is tan (el), the azimuth error resulting from axis skewness is

$$\Delta Az = \xi \tan (el) \tag{14}$$

or, in terms of XEl,

$$\Delta XEl = \xi \sin (el) \tag{15}$$

The elevation error due to skewness is

$$\Delta el = \tan (el') - \tan (el) = \frac{X_2\xi}{X_1} \tag{16}$$

which becomes

$$\Delta el = \xi \tan (\Delta Az) \tag{17}$$

Since the azimuth skewness is small, the skewness error in elevation will be small and can therefore be neglected.

Skew terms for a ha/dec-type mount are again defined in the same manner detailed above, with hour angle and declination replacing azimuth and elevation, respectively.

C. Azimuth (Hour Angle) Axis Tilt

Axis tilt is the deviation of the azimuth from the true vertical. For a polar mount, axis tilt is the deviation of the hour angle axis from the north celestial pole, or the amount of nonparallelism with the earth's spin axis. Figure 5 depicts X_1 as the neutral axis with ϕ being the angle at which the azimuth plane is tilted.

The azimuth error can be found by determining the rate of change of Q_K (which is the azimuth of the level axis (K_A) minus azimuth) with respect to the change in the magnitude of the X_2 coordinate

$$\Delta Az = \frac{d(K_A - Az)}{dX_2} \Delta X_2$$

$$= \frac{d\left[\tan^{-1}\left(\dfrac{X_2}{X_1}\right)\right]}{dX_2} \Delta X_2 \tag{18}$$

where $\Delta X_2 = X_2' - X_2$. The result of performing the differentiation is

$$\Delta Az = \frac{-X_1 d(X_2)}{[(X_2)^2 + (X_1)^2] dX_2} (X_2' - X_2) \qquad (19)$$

The matrix to transform from an untilted coordinate system to a coordinate system tilted by an angle ϕ as given by $[x'] = [a][x]$ is

$$[A] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \phi \\ 0 & -\phi & 1 \end{bmatrix} \qquad (20)$$

if ϕ is small. One result of the matrix transformation is

$$X_2' - X_2 = (X_2 + X_3\phi) - X_2 = \phi X_3 \qquad (21)$$

which changes Eq. (19) to

$$\Delta Az = \frac{-X_1 d(X_2)}{[(X_2)^2 + (X_1)^2] dX_2} \phi X_3$$

$$= \frac{[\sin(K_A - Az)] \phi\rho \sin(el)}{\rho \cos(el)}$$

$$= -\phi \tan(el) \sin(K_A - Az) \qquad (22)$$

The elevation error resulting from azimuth plane tilt is

$$\Delta el \cong \tan(el) - \tan(el') \qquad (23)$$

Using the transformation matrix and the fact that

$$\tan(el) = \frac{X_3}{\rho \cos(el)} \qquad (24)$$

$$\tan(el') = \frac{X_3'}{\rho \cos(el)} \qquad (25)$$

then

$$\Delta el = \frac{X_3}{\rho \cos(el)} - \frac{-\phi X_2 + X_3}{\rho \cos(el)} = \frac{\phi X_2}{\rho \cos(el)} \qquad (26)$$

$$\Delta el = \phi \cos(K_A - Az)$$

Tilt terms for a ha/dec-type mount are also defined in the same manner detailed above, with hour angle and declination replacing azimuth and elevation, respectively.

D. Structural Flexure Due to Gravity

Forces due to gravity result in moments upon the structure, deflecting the truss members and panels, thus changing the shape of the antenna. These forces and moments change directions and magnitude with the elevation angle. Due to the gravitational deformation of the structure, the antenna beam is deflected relative to the shaft encoder on the elevation axis.

The antenna structural error under gravity loading is computed as the algebraic sum of the individual errors, each caused by a single structure component under one type of deformation (e.g., translation and rotation). The three key structural components considered are the main reflector, the subreflector, and the feed cone. The five individual errors of interest as shown in Fig. 6 are main reflector axis rotation, main reflector vertex translation, subreflector vertex translation, subreflector axis rotation, and feed translation.

Each element of deformation could generate a boresight pointing error in space about the x-axis ($\Delta\theta_x$) and about the gravity y-axis ($\Delta\theta_y$). Due to the almost symmetrical antenna geometry about the y-z plane $\Delta\theta_y$ is negligible and only the pointing errors $\Delta\theta_x$ are considered. The total pointing error under gravity loading about the x-axis is expressed as the

$$\sum_j \theta_x(j)$$

Geometrical factors are introduced for each type of deformation to translate the component into an antenna rf beam error (Ref. 6).

Each error component $\Delta\theta_x(j)$ can be represented as the resultant effect of pointing errors at the true extreme elevation positions:

$$\Delta\theta_x(j) = \Delta\theta_x^y(j)[\cos(el) - \cos(\alpha_r)]$$

$$+ \Delta\theta_x^z(j)[\sin(el) - \sin(\alpha_r)] \qquad (27)$$

where α_r is the antenna-panel setting (rigging) elevation angle (usually 45 degrees), $\theta_x^y(j)$ is the rotation about the x-axis due to weight loading applied parallel to the reflector y-axis and $\theta_x^z(j)$ is the rotation about the x-axis due to weight loading applied parallel to the reflector z-axis (Refs. 7 and 8). These two loading conditions provide sufficient information to compute the structural pointing error at any elevation angle of the antenna.

For a ha/dec-type mount the *az/el* expression for flexure given in Eq. (27) is converted to ha/dec coordinates using the identities of spherical trigonometry, resulting in (Refs. 9 and 10)

$$\Delta HA = -\sin(ZD)\sin(\beta) - X'_c \qquad (28)$$

$$\Delta DEC = -\sin(ZD)\cos(\beta) - Y'_c \qquad (29)$$

where ZD is the zenith angle (90 degrees $- el$), β is the paralectric angle, X'_c is the component value of the surface panels setting angle and Y'_c is the component value of the surface panel declination angle. The pointing correction in terms of *xdec* is

$$\Delta xdec = [-\sin(ZD)\sin(\beta) - X'_c]\cos(dec) \qquad (30)$$

Other flexure corrections are included to compensate for subreflector position error. Due to deformation of the main reflector, the focii of the main reflector and subreflector do not coincide. The subreflector offset (in inches) required to accommodate this condition is

$$y - \text{offset} = 3.41[\sqrt{2}\cos(el) - 1] \qquad (31)$$

and is part of the Subreflector Controller (SRC) software. A correction is also made for the boresight shift due to the lateral motion of the subreflector, "squint" correction, is

$$\text{"squint" correction} = (y - \text{offset})\frac{(M-1)R}{MF} \qquad (32)$$

where

M = magnification factor of the subreflector

R = reflection ratio

F = focal length

The squint correction algorithm is located in the ACS.

E. Residual Error Due to Refraction

The residual pointing error is the resultant error that remains after the "standard" corrections (Ref. 1) for bending are made in the refraction correction algorithm of the ACS. At the present time, the correction for the residual pointing error is made using the following expression

$$\Delta el = K\cot(el) \qquad (33)$$

where K is the constant that takes into account the residual effects of temperature, pressure, and humidity.

IV. Pointing Model Generation

A pointing model is used to characterize the various sources of pointing error for the antenna and consists of the items detailed in the previous section. An az-el antenna will be used in this section to demonstrate the methodology involved in the usage and development of new models.

The equation for the elevation and cross-elevation pointing offsets are

$$\Delta Xel = P1 + P2\cos(el)$$
$$+ P3\sin(el)$$
$$+ P4\sin(el)\cos(az)$$
$$+ P5\sin(el)\sin(az) \qquad (34)$$

$$\Delta el = -P4\sin(az)$$
$$+ P5\cos(az) + P7$$
$$+ P8\cos(el) + P9\cot(el) \qquad (35)$$

where

$P1 = \delta$, the magnitude of collimation axis misalignment

$P2$ = azimuth encoder fixed offset

$P3 = \xi$, elevation axis skew angle

$P4 = -\phi\sin K_A$ (ϕ is the azimuth axis offset and K_A is the azimuth of the level axis)

$P5 = \phi\cos K_A$

$P7$ = elevation encoder fixed offset and elevation collimation error

$P8 = \Delta\theta^y_x$, the antisymmetry loading pointing error computed for a hypothetical horizon position ideal cassegrainian antenna under unit gravity off/on loading

$P9 = K$ (residual refraction error)

Initially, pointing offset data were collected for various conscan tracks using a program called CAPTURE. Approximately every two seconds, this program reads the azimuth and elevation angles, azimuth and elevation offsets, and time tags the data. The azimuth offset is next converted to *XEl* coordinates, so it can be entered into the model fitting program (acronym PHO). The PHO program uses a linear least-squares technique to fit the offset data to the functions given in Eqs. (34) and (35). The least squares fit determines the parameters P_i. The systematic error correction table can then be generated and loaded into the Antenna Pointing Assembly

(APA) and then into the ACS. These past efforts resulted in sets of parameters to be used for specific tracks.

V. Conclusion

Figures 7 and 8 demonstrate the effect of model corrections by showing a set of offset data before and after corrections are made. The plots show the offsets after correction are within 10 percent of the half-power beamwidth.

So far, the pointing effort is concentrating on improving and verifying the pointing model and adjusting those initially developed parameter sets. The measureable parameters (i.e., tilt, orthogonality, etc.) of Eqs. (34) and (35) will be field

verified and a Fourier analysis will be performed to ensure that no more systematic errors are present and only random errors remain. When it is determined that the selected set of parameters are acceptable and all of the systematic errors seem to be accounted for, alternate methods of fitting the offset data to the functions of Eqs. (34) and (35) will be examined.

Secondly, since conscan data sets were intially collected and identified with specific tracks, many systematic error correction models are presently being used. It is believed that these models can be condensed into one model representing error corrections over the entire sky. Finally, a technique will be studied to update these models in a timely fashion as new offset data are collected.

References

1. Berman, A. L., and Rockwell, S. T., "New Optical and Radio Frequency Angular Tropospheric Refraction Models for Deep Space Applications," *JPL Technical Report 32-1601*, Jet Propulsion Laboratory, Pasadena, Calif., November 1975.

2. Ohlson, J. E., and Reid, M. S., "Conical Scan Tracking with the 64 m Diameter Antenna at Goldstone," *NASA Technical Report 32-3605*, Jet Propulsion Laboratory, Pasadena, Calif., October 1976.

3. Meeks, M. L., Ball, J. A., and Hull, A. B., "The Pointing Calibration of the Haystack Antenna," *IEEE Trans. on Ants. and Prop., AP-16*, No. 6, pp. 746–751, November 1968.

4. Burak. M., "Pointing Calibration of a High-Resolution Millimeter-Wave Antenna by Star Observations," *Air Force Systems Command AFCRL-71-0220*, Air Force Cambridge Research Laboratories, March 1971.

5. Stumpff, P., "Astronomical Pointing Theory for Radio Telescopes," *Klein-heubacher Berichte, 15*, pp. 431–437, 1972.

6. Isber, A. M., "Obtaining Beam-Pointing Accuracy With Cassegrain Antennas," *Microwaves, 6*, No. 8, pp. 40–44, August 1967.

7. Katow, M. S., and Mori, M., "Computation of RF Boresight Direction from Reflector Distortions," *JPL Technical Report 32-1526, XVII*, pp. 78–82, Jet Propulsion Laboratory, Pasadena, Calif. 1973.

8. Katow, M. S., "64 m Diameter Antenna: Computations of RF Boresight Direction," *JPL Technical Report 32-1526, XIV*, pp. 68–72, Jet Propulsion Laboratory, Pasadena, Calif., 1973

9. Katow, M. S., and Levy, R., "Computation of Gravity RMS for HA-DEC Antennas," *DSN Progress Report 42-27*, pp. 139–147, Jet Propulsion Laboratory, Pasadena, Calif., 1975.

10. Katow, M. S., "34-Meter Antenna-Subreflector Translations to Maximize RF Gain," *TDA Progress Report 42-62*, pp. 112–120, Jet Propulsion Laboratory, Pasadena, Calif., 1981.
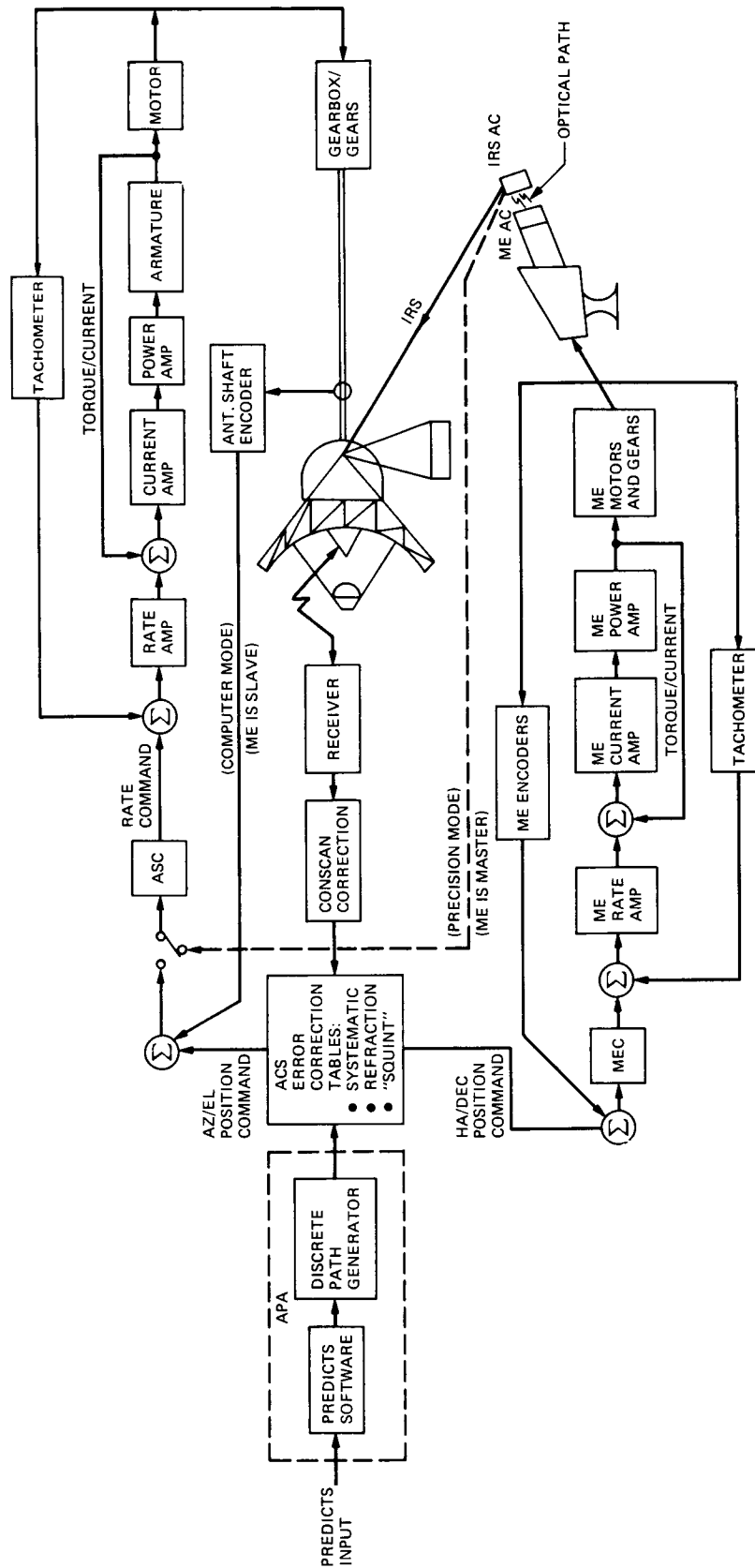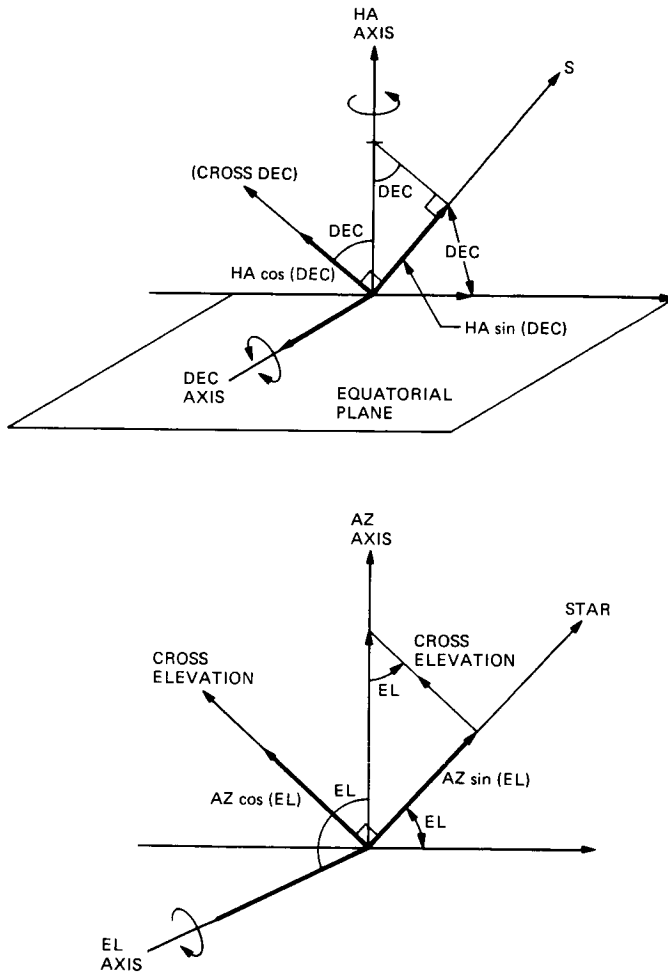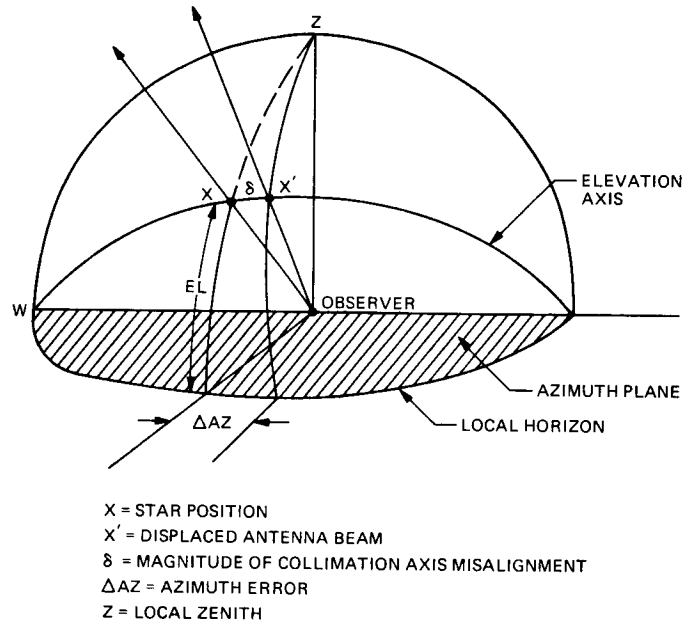
Fig. 1. Overview of the antenna pointing system

Fig. 2. Azimuth-elevation and hour-angle declination coordinate systems showing cross elevation and cross-declination axes, respectively
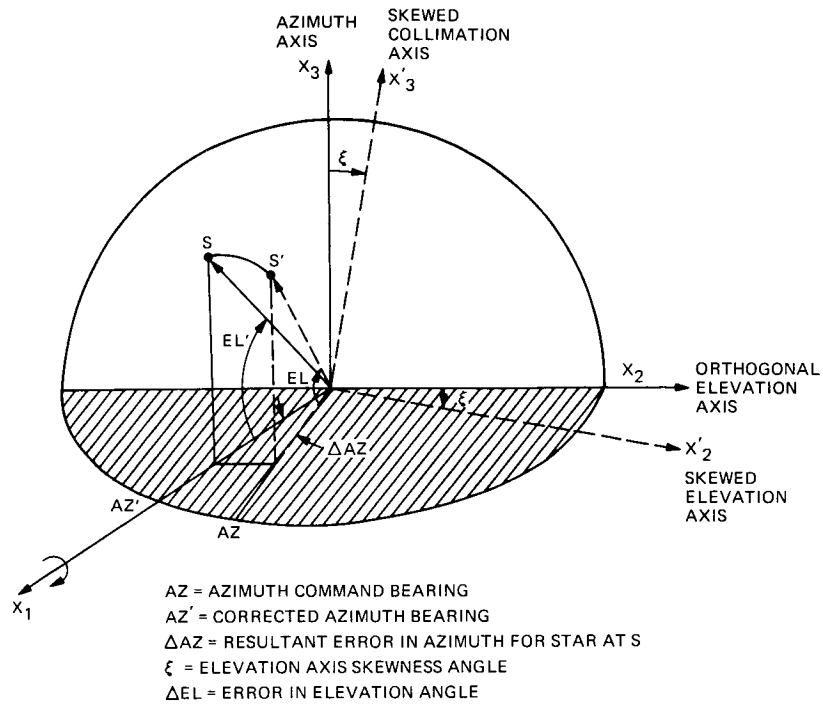

X = STAR POSITION
X' = DISPLACED ANTENNA BEAM
δ = MAGNITUDE OF COLLIMATION AXIS MISALIGNMENT
ΔAZ = AZIMUTH ERROR
Z = LOCAL ZENITH

Fig. 3. Azimuth collimation error

AZ = AZIMUTH COMMAND BEARING
AZ' = CORRECTED AZIMUTH BEARING
ΔAZ = RESULTANT ERROR IN AZIMUTH FOR STAR AT S
ξ = ELEVATION AXIS SKEWNESS ANGLE
ΔEL = ERROR IN ELEVATION ANGLE

Fig. 4. Elevation axis skewness

S = STAR POSITION
S' = DISPLACED ANTENNA BEAM
ϕ = MAGNITUDE OF AZIMUTH PLANE TILT ANGLE
K_A = AZIMUTH OF LEVEL AXIS
Q_K = AZ - K_A
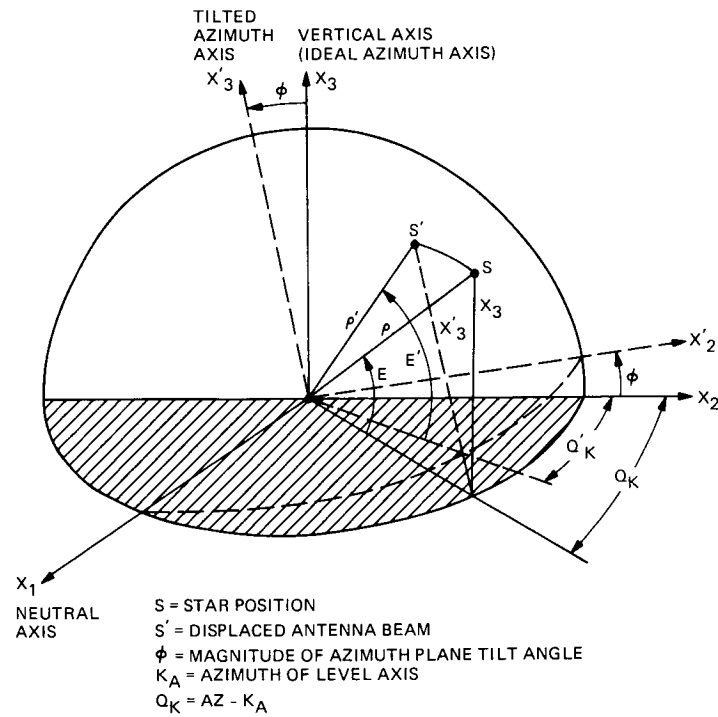
Fig. 5. Azimuth plane tilt
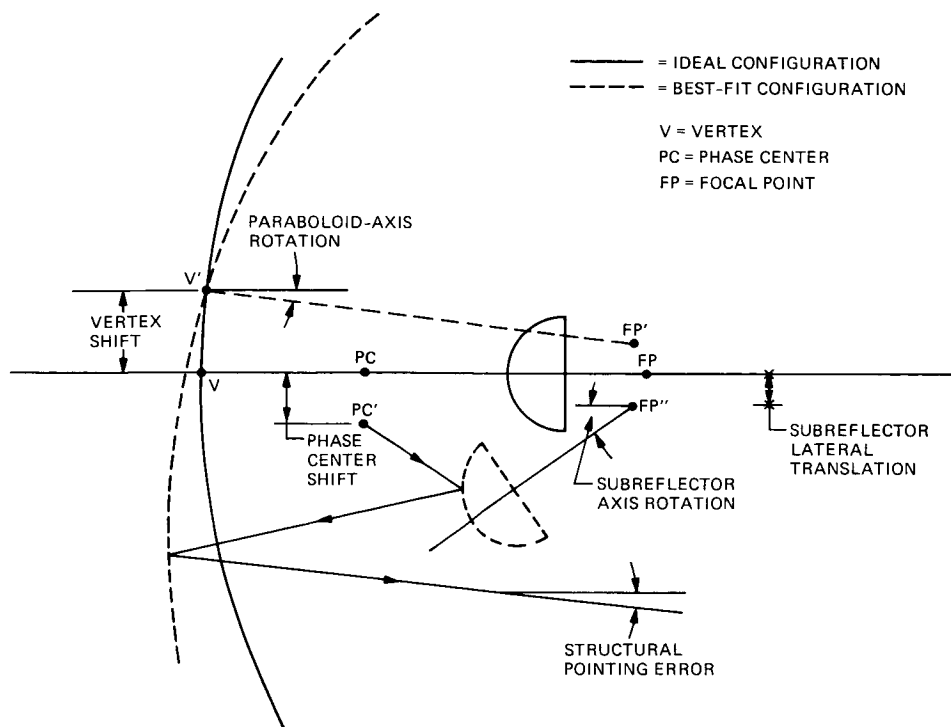
Fig. 6. Structural feed misalignments
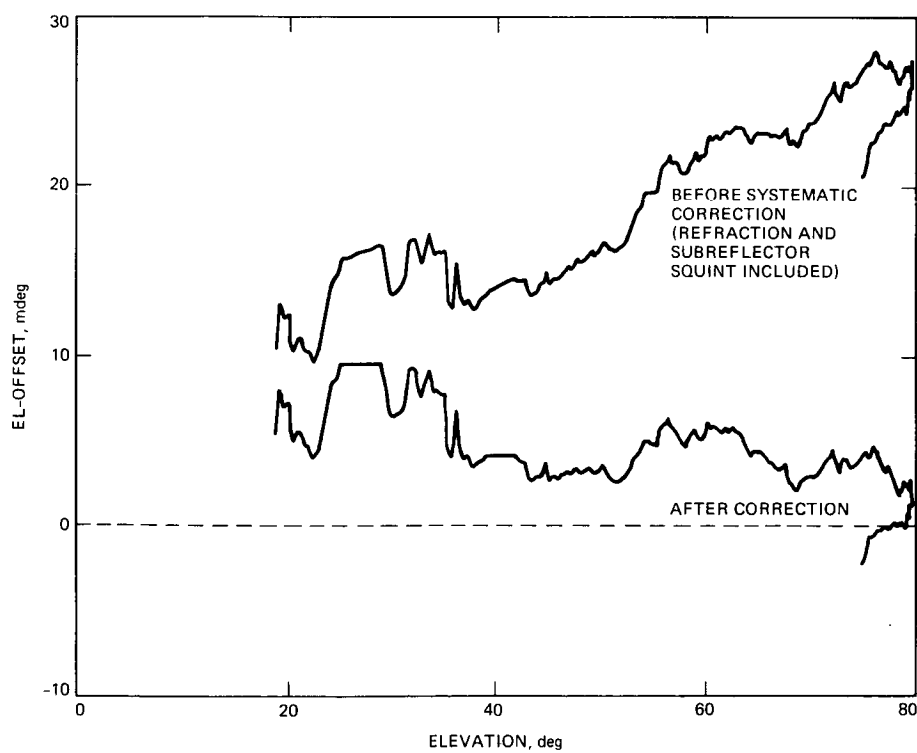
Fig. 7. Elevation vs *EL*-offset before and after systematic error correction, Mars, 64-m antenna
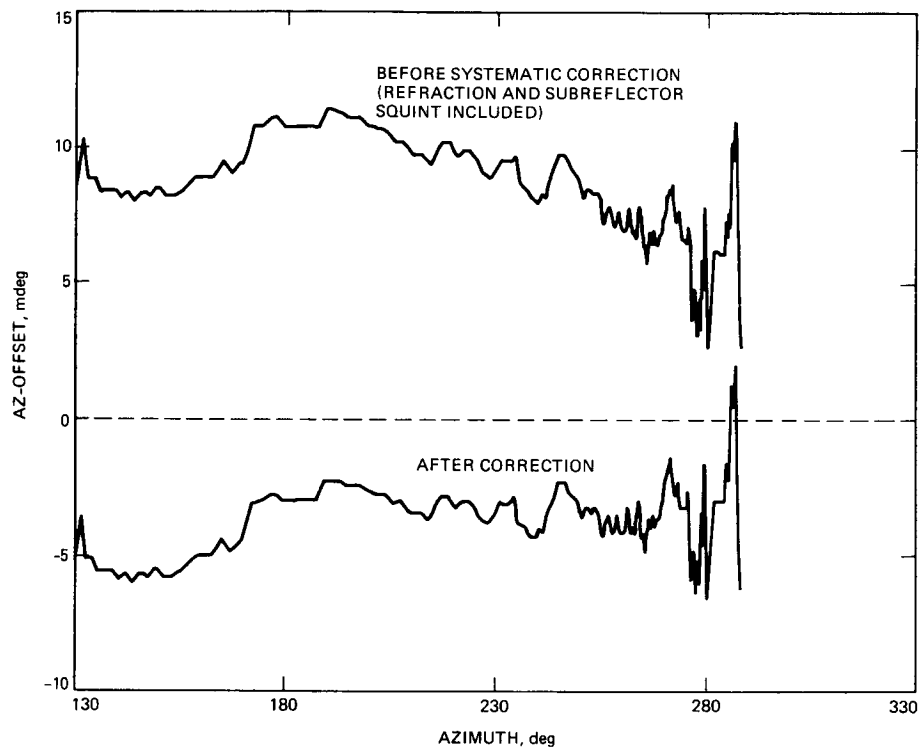
Fig. 8. Azimuth vs *AZ*-offset before and after systematic error correction, Mars, 64-m antenna

Erasure Declaring Viterbi Decoders

F. Pollara
Communications Systems Research Section

Several methods for realizing erasure declaring Viterbi decoders for the (7, 1/2) NASA code are discussed. Only bit oriented algorithms are considered. When such decoders are used in a concatenated system with a (255,223) Reed-Solomon decoder, improvements on the probability of word error of at most 0.1 dB were obtained.

I. Introduction

Reliable deep space communication can be realized with concatenated coding systems based on an inner convolutional code and an outer Reed-Solomon code. A (7, 1/2) inner convolutional code and an 8-bit (255,223) outer Reed-Solomon code are used in NASA's Voyager mission and as an international coding standard.

This Reed-Solomon code can correct any word such that $2e + E \leqslant 2t$, where e is the number of symbol errors in the word, E the number of erasures, and $2t = 32$ the number of parity symbols. While Reed-Solomon decoders which can correct erasures can be easily implemented (Ref. 1), methods for estimating symbol quality and criteria for declaring symbol erasures are open to discussion.

There are two distinct classes of methods: one extracts quality information from the Viterbi decoder, the other examines adjacent symbols of interleaved Reed-Solomon words and erases symbols next to incorrect symbols. Both these classes of methods have been considered in Ref. 2. We will concentrate on further variations of the first class methods, which give better results than those reported in Ref. 2, and comparable results to methods belonging to the second class.

II. Reed-Solomon Code Performance

Our (255,223) Reed-Solomon code over $GF(2^8)$ has a word error probability given by Ref. 3:

$$P_w = \sum_{i=0}^{n} \sum_{\substack{j=2(t-i)+1 \\ j \geqslant 0}}^{n-i} \alpha p^i s^j (1 - p - s)^{n-i-j} \qquad (1)$$

where

$$\alpha = \frac{n!}{i! j! (n - i - j)!}$$

and $n = 255$; p is the symbol error rate at the Reed-Solomon decoder input; and s is the symbol erasure rate. The only ingredients needed to compute the performance of the concatenated system are therefore p and s, which are measured at the output of a software simulated Viterbi decoder, driven by convolutionally encoded data in additive Gaussian noise. The values of E_b/N_0 shown in this report have been increased by 0.58 dB to take into account the Reed-Solomon code rate. These values represent the correct SNR of the concatenated channel.

III. Erasure Declaring Viterbi Decoders

The general problem is that of modifying the Viterbi algorithm so that an estimate on the quality of decoded bits or symbols (8 bits in our case) can be provided.

Theoretically we should be able to compute the a posteriori probability for each bit or symbol and then compare it to a threshold to declare erasures. Since the Reed-Solomon decoder operates on symbols, we are ultimately interested in the a posteriori symbol probability. Lee (Ref. 4) devised a decoding algorithm called real-time minimal byte error probability (RTMBEP) decoder, which actually provides the a posteriori symbol probability, but is unfortunately too complex to implement. A similar MAP bit decoding algorithm has been proposed by P. L. McAdam (Ref. 5), but suffers from similar implementation complexity problems.

Byte-oriented decoders should offer better erasure information on Reed-Solomon symbols, but due to their complexity we restrict ourselves to conventional bit-oriented Viterbi decoders.

Several modified Viterbi algorithms will be described as follows.

Method A. Extraction of quality information for Viterbi decoded bits from the rate of metric renormalization has been considered in Ref. 2, and will not be repeated here. This method's performance is severely limited by the fact that the renormalization rate cannot resolve precisely enough in time which bits are less reliable.

Method B. The conventional Viterbi algorithm searches all possible code sequences (paths) and finds the most likely transmitted sequence. Accumulated metrics can be viewed as distances between surviving paths and the received sequence, where the closest path has metric normalized to zero and all other paths have some positive metric. These metrics do not contain enough information to reconstruct the a posteriori probabilities of paths, since some paths are pruned at each stage and their effect is thereafter ignored, together with the probabilities of all the paths that could have departed from them. Yet the accumulated metrics do contain "some" information about the reliability of each surviving path. Typically the metric values will be spread over a wide range of positive values up to approximately $2(K - 1)$ (Ref. 6), where K is the constraint length of the code, if SNR $\gg$ 1, while they will tend to accumulate around zero if SNR is low. Therefore the reliability of a chosen path (zero metric) can be estimated according to the *spread* of the metrics.

Another basic problem is that we are interested in the quality of a decoded bit and not in that of an entire path. Different paths may or may not yield identical decoded bits.

Once the quality of a bit has been decided, we will declare erasure by comparing it with a threshold. Ultimately, we will have to decide on symbol (byte) erasures, which can be declared if one or more bits are erased in the symbol. Although this study is limited to bit-oriented Viterbi decoders, symbol erasure criteria based on multiple bits erasures or multiple bit cumulative quality have been tested. No measurable improvements were found as compared to declaring symbol erasure based on a single bit erasure in the symbol.

Let m_j be the accumulated metric of state j at a given time t; $b_j \in (+1, -1)$ be the information bit belonging to the surviving sequence into state j at time $t - \tau$; and j^* be the state such that $m_{j^*} = 0$, i.e., the state chosen as most likely at time t, as shown in Fig. 1. Then b_{j^*} will be the decoded bit at time t.

In a Viterbi decoder with $M = 2^{K-1}$ states, Method B forms the sum,

$$B = \sum_{j=0}^{M-1} b_j, \qquad b_j \in (-1, +1)$$

and then declares erasure if $Bb_{j^*} < 0$. This corresponds to a simple majority rule, where an erasure is declared if less than half of the b_j agree with b_{j^*}.

Method C. Let

$$m_{max} = \operatorname*{Max}_{j} \{m_j\}$$

and $m'_j = m_{max} - m_j, j = 0, \ldots, M - 1$. This method defines B as

$$B = \sum_{j=0}^{M-1} m'_j b_j$$

and declares erasure if

$$Bb_{j^*} < -\frac{1}{2} \sum_{j=0}^{M-1} m'_j$$

This corresponds to a weighted majority rule, where a higher weight is assigned to more reliable paths.

Method D. This method (proposed by D. Divsalar) is substantially different from all others. It stores survivors as strings of symbols from a ternary alphabet $(+1, -1, E)$, where E represents erasure and $b_j \in (+1, -1, E)$. An erasure is stored when the relative difference of the two competing accumulated metrics is less than a threshold T. Let these two metrics be denoted by $m_j^{(a)}$ and $m_j^{(b)}$, where

$$m_j = \min_j \{m_j^{(a)}, m_j^{(b)}\}$$

then an erasure is stored if

$$\frac{|m_j^{(a)} - m_j^{(b)}|}{m_j^{(a)} + m_j^{(b)}} < T$$

where T is small constant. An erasure is declared if $b_{j*} = E$.

Method E. This method defines B as

$$B = \sum_{j=0}^{M-1} e^{-m_j} b_j$$

and declares erasure if

$$B b_{j*} < T \sum_{j=0}^{M-1} e^{-m_j},$$

where T is a threshold value. This corresponds to a nonlinearly weighted majority rule.

Method F. Let J_1 and J_0 be two sets of values of j such that $b_j = 1$ and $b_j = 0$, respectively. Then, this method defines B_1 and B_0 as

$$B_1 = \sum_{j \in J_1} e^{-m_j}$$

$$B_0 = \sum_{j \in J_0} e^{-m_j}$$

and declares erasure if

$$b_{j*} \log\left(\frac{B_1}{B_0}\right) < T,$$

which is equivalent to a log-likelihood ratio test.

IV. Results and Conclusion

All the above methods have been tested extensively by simulation, for various threshold values. Thresholds showed a mild dependency on E_b/N_0, and have been optimized by repeated trials.

Symbols containing at least one reliable bit have been erased. Erasing only symbols with at least two unreliable bits has been tried and proved inferior.

Among all methods described, Method F performed consistently, though slightly, better than any other. The superiority of Method F is due to the fact that this method yields the closest possible approximation of true a posteriori probabilities, based on the Viterbi algorithm. The performance of this method in terms of probability of word error at the output of Reed-Solomon decoder (Eq. [1]) vs. concatenated E_b/N_0 is shown in Fig. 2, where the performance of the usual (no erasure) system is also shown for comparison. The curve denoted as "lower bound" is the performance of an hypothetical system, where a "genie" knows exactly all symbols in error, which are then erased.[1] Aside from the lower bound, the two sets of curves in Fig. 2 represent two different truncation lengths L of survivors in the Viterbi decoder.

These simulation results indicate that an improvement of approximately 0.1 dB can be obtained at $P_w = 10^{-5}$ by an erasure declaring Viterbi decoder, based on Method F, if $L = 32$ bits. This result compares favorably with that obtained in Ref. 2. However, the improvement becomes negligible if the truncation length is increased to 64 bits.

These results seem to suggest that the additional gain available may be achievable only with the more complex, symbol oriented, algorithms based on true a posteriori symbol probabilities.

[1] This method was devised by L. Deutsch, and reported in JPL IOM 331-83-132A (internal document), by L. Swanson, Jet Propulsion Laboratory, Pasadena, Calif., April 12, 1983.

References

1. Blahut, R. E., *Theory and Practice of Error Control Codes*. New York: Addison-Wesley, 1983.

2. Pitt, G. H., III and Swanson, L., "Erasure Information for a Reed-Solomon Decoder," *TDA Progress Report 42-83*, pp. 39–44, Jet Propulsion Laboratory, Pasadena, Calif., November 1985.

3. Clark, G. C., and Cain, J. B., *Error-Correction Coding for Digital Communications*. New York: Plenum Press, 1981.

4. Lee, L. N., "Concatenated Coding Systems Employing a Unit-Memory Convolutional Code and a Byte-Oriented Decoding Algorithm," *IEEE Trans. on Communications*, Vol. COM–25, No. 10, pp. 1064–74, October 1977.

5. McAdam, P. L., *MAP Bit Decoding of Convolutional Codes*, Ph.D. Disseration, University of Southern California, Los Angeles, February 1974.

6. Viterbi, A. J., and Omura, J., *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
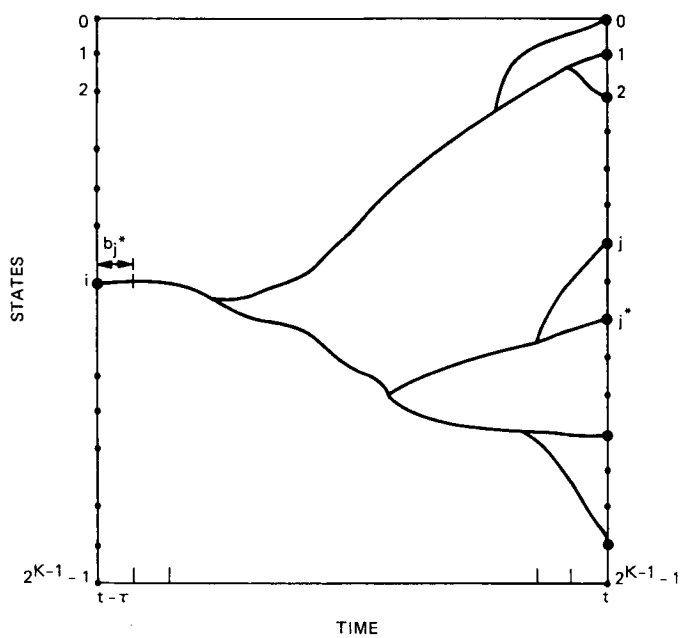
Fig. 1. Example of surviving sequences at time t
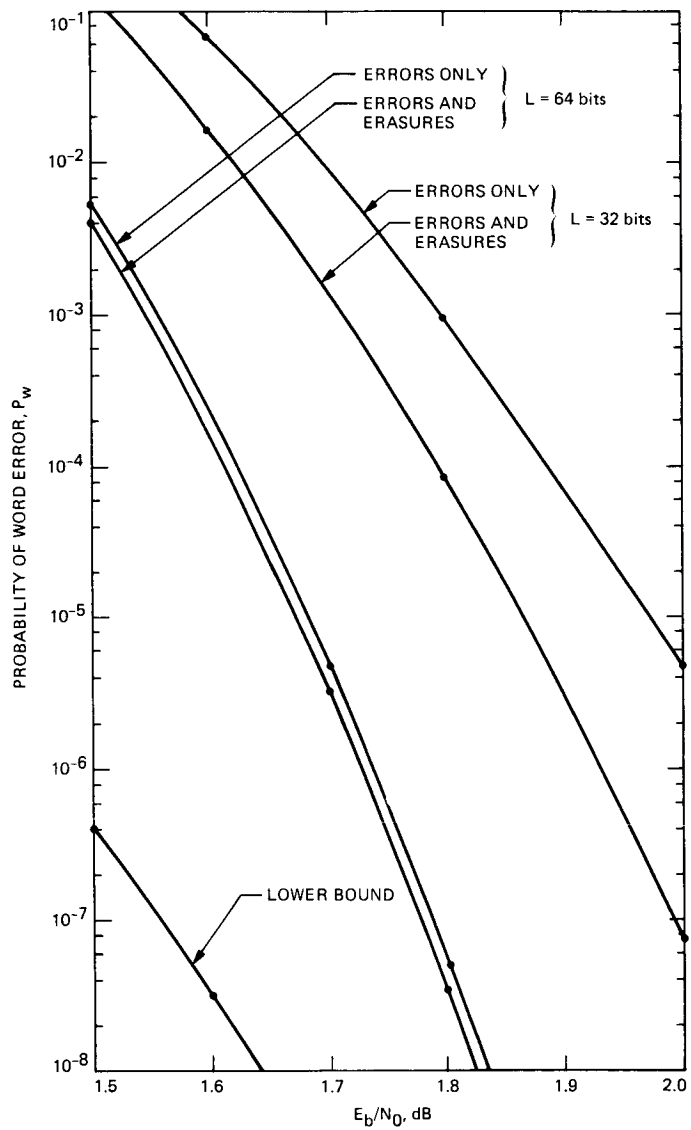
Fig. 2. Probability of word error vs E_b/N_0 for Method F

A VLSI Architecture of a Binary Updown Counter

I. S. Reed

University of Southern California

I. S. Hsu and T. K. Truong

Communications Systems Research Section

In this article, a new pipeline binary updown counter with many bits is developed which can be used in a variety of applications. One such application includes the design of a digital correlator for very long baseline interferometry (VLBI) [1].

The advantage of the presently conceived approach over the previous techniques is that the number of logic operations involved in the design of the new binary updown counter can be reduced substantially. The architecture design using these methods is regular, simple, expandable and, therefore, naturally suitable for VLSI implementation.

I. Introduction

Large binary updown counters are widely used in digital circuits. Examples include the controller design of a servomotor or the correlator part of very long baseline interferometry (VLBI) [1]. In [7], a VLSI design of an up counter is used as a model for an updown counter. In that design, only two one-bit registers and a half adder are needed in each stage and its expansion to implement a counter with more bits is very easy and straight forward. However, a fixed bias value is added to the sum. Therefore, the sum "zero" is represented as the middle value of the range of the numbers and the smallest negative number is represented by "zero". Also, it takes n clock times to perform one counting operation in the worst case.

In this article, a new design of a large pipeline binary updown counter is presented. The design of this new updown

counter is simple, regular and expandable. Also, only one clock time is needed to perform one counting operation. Using this new architecture it is verified that a 64-bit binary updown counter can be put readily on a single VLSI chip with current NMOS technology. An example describing the pipeline architecture as well as the simplicity of each basic cell of this new counter is illustrated completely for a 3-bit case.

II. A New Algorithm for a Binary Updown Counter

In this section, a new algorithm is developed for the implementation of a binary updown counter. This algorithm is illustrated by an example for a 3-bit case. However, the same structure clearly can be extended to the more general cases.

Let the input to this counter have three values, namely, -1, 0, and 1. The input is fed into the counter sequentially. Also let $A_n(t)$ denote the value of the nth bit of a clocked counter at time t. Then it can be verified easily by the techniques in [2] that

$$A_n(t) = A_n(t-1) \oplus ((U \cdot P_n) \cup (D \cdot Q_n)) \qquad (1)$$

where there is no loss of generality to assume time t is integer-valued and where $A_n(t-1)$ is the value of the nth bit of an updown counter at time $t-1$. Also in Eq. (1) the symbol "$\oplus$" denotes "exclusive OR" operation. The symbol "$\cdot$" denotes "AND" operation and the symbol "$\cup$" is the "inclusive OR" operation. If $U = 0$ and $D = 1$, then the input is "-1". If both U and D equal zero, then the input is "0". In Eq. (1), P_n represents product of A_k's, for $1 \leqslant k \leqslant n-1$, while Q_n is the product of $\overline{A_k}$'s, for $1 \leqslant k \leqslant n-1$, where $\overline{A_k}$ is the complemented value of A_k. Both P_n and Q_n are expressed as the following two equations:

$$P_n = A_{n-1} A_{n-2} \cdots A_1 \qquad (2\text{-a})$$

$$Q_n = \overline{A}_{n-1} \overline{A}_{n-2} \cdots \overline{A}_1 \qquad (2\text{-b})$$

Equation (2) can be rewritten in a recursive form as

$$P_n = P_{n-1} \cdot A_{n-1} \qquad (3\text{-a})$$

and

$$Q_n = Q_{n-1} \cdot \overline{A}_{n-1} \qquad (3\text{-b})$$

with the initial values $P_1 = 1$ and $Q_1 = 1$. Intuitively, Eq. (1) can be described as follows:

If the input to the counter is a "0", then the contents of the counter should remain unchanged. It is easily verified from Eq. (1) that $A_n(t) = A_n(t-1)$ with $U = D = 0$.

If the input is "0", both U and D equal "0" according to the previous assignments. This will null the second term in Eq. (1) and by the property of "exclusive OR", one yields $A_n(t) = A_n(t-1)$.

On the other hand, if the input is "1", then $U = 1, D = 0$. The second term of Eq. (1) $(U \cdot P_n \cup D \cdot Q_n)$, is then changed to $(1 \cdot P_n) \cup (0 \cdot Q_n) = P_n$. This equals 1 if $P_n = 1$ which implies $A_1 = A_2 = \cdots = A_{n-1}$. In counting up, the value of nth bit changes from zero to one if (1) input to the counter is a "1" and (2) the value of bit 1 to bit $n-1$ are all ones. This

fact is truthfully reflected by Eq. (1). A similar argument applies to the case when the input to the counter is a "-1", i.e., the counter is counting down.

III. Example of the New Updown Counter

We illustrate the new algorithm with the design of a 3-bit binary updown counter.

The truth table of a 3-bit binary updown counter is shown in Table 1. Where $A_i(t)$, for $i = 1, 2, 3$, denotes value of ith bit of the counter at time t. To verify this truth table, consider, for illustation the 4th row in Table 1.

In this row,

$$D = 0, \quad U = 1, \quad A_3(t) = 0, \quad A_2(t) = 1, \quad A_1(t) = 1.$$

This corresponds to the case for which the contents of counter is digit 3 in binary and the input is "1". Obviously, the next value should equal four. Namely, $A_3(t+1) = 1, A_2(t+1) = 0$ and $A_1(t+1) = 0$.

Substituting these values of A_i, for $i = 1, 2, 3$, as well as the values $U = 1, D = 0$ into Eq. (1), yields the following results:

$$A_3(t+1) = A_3(t) \oplus (U \cdot P_3 \cup D \cdot Q_3)$$

$$= 0 \oplus (1 \cdot 1 \cup 0 \cdot 0)$$

$$= 0 \oplus 1$$

$$= 1$$

In the above substitutions one uses the fact that

$$P_3 = A_1 \cdot A_2 = 1$$

and

$$Q_3 = \overline{A}_1 \cdot \overline{A}_2 = 0.$$

Also,

$$A_2(t+1) = A_2(t) \oplus ((U \cdot P_2) \cup (D \cdot Q_2))$$

$$= 1 \oplus ((1 \cdot 1) \cup (0 \cdot 0))$$

$$= 1 \oplus 1$$

$$= 0$$

and

$$A_1(t+1) = A_1(t) \oplus ((U \cdot P_1) \cup (D \cdot Q_1))$$

$$= 1 \oplus ((1 \cdot 1) \cup (0 \cdot 0))$$

$$= 1 \oplus 1$$

$$= 0$$

The rest-states of Table 1 can be verified in a similar manner.

IV. A VLSI Architecture for Implementing an n-bit Binary Updown Counter

In this section, a VLSI architecture is developed for an n-bit binary updown counter. This VLSI counter is composed of n basic cells as well as a data mapping programmable logic array (PLA) [3]. The ith basic cell carries out the operations for calculating P_i, Q_i, and A_i.

Figure 1 shows a pipeline architecture for an n-bit binary updown counter. It is composed of n basic cells and a data conversion PLA. Input to the data conversion PLA is either 1, 0 or –1. These three values are represented as 01, 00 and 10 respectively in binary representation. It is true that if the input numbers are represented in accordance with this format, then this PLA is not needed. However, it is required in general to match to the outside system. Output of this PLA are values of U and D. As described in the previous section, if input value is a "1" then $U = 1$ and $D = 0$. If input is a "–1", then $U = 0$ and $D = 1$. Both U and D are zero if input is a "0". In Fig. 2, a block diagram as well as a table illustrate the relationship between the input and output of this PLA.

Inputs to the ith basic cell are U, D, P_i and Q_i while the outputs are U, D, P_{i+1}, Q_{i+1} and A_{i+1}. Outputs P_{i+1} and Q_{i+1} are obtained by the calculation of Eqs. (2a) and (2b). A_{i+1} is the intermediate value of a switch bit of this counter.

For an n-bit binary updown counter, n identical basic cells are required in the design.

Figure 3 shows the logic diagram of a basic cell in Fig. 1. This basic cell consists of three one-bit shift registers where registers labeled as RQ_i and RP_i are used to store values of Q_i and P_i respectively. Register RA_i stores value of A_i. The output of RA_i is sent to an XOR circuit for calculating the value of $A_i(t+1)$. In the mean time, A_i together with its complementary value $\bar{A}_i$ is sent to two AND gates for the calculation of P_{i+1} and Q_{i+1}. Figure 4 shows the logic diagram of a one-bit shift register. The two-phase clocking scheme is adopted in this design for the ease of timing control.

As was described previously, the initial values of P_n and Q_n are $P_1 = Q_1 = 1$. Therefore the inputs to registers of the first basic cell RP_1 and RQ_1 are tied to VDD which is always at logic value "one".

Figure 5 shows the pin assignment of an n-bit binary updown counter. Where VDD and GND are power pins, ϕ_1 and ϕ_2 are two inputs for the non-overlapping clocks. A_i, for $1 \leqslant i \leqslant n$ are the output pins representing the results of counting. The final value is obtained n time units after the last input fed into the counter. Figure 6 shows the layout of a 64-bit binary updown counter. This was carried out by using the CAD system described in [8]. The total chip area is approximately $6000 \times 5700 \ \mu m^2$. The area occupied by the counter circuit is only $2800 \times 2600 \ \mu m^2$. The difference is due to the large number of input and output pads needed to interface with the outside world.

V. Conclusion

An efficient architecture has been developed for the VLSI implementation of a binary updown counter. The architecture is easily extensible in the number of bits counted. A 64-bit binary updown counter is designed according to this new algorithm. It is demonstrated in this article that a 64-bit binary updown counter can be realized easily on a single VLSI chip with current NMOS technology.

References

1. Thomas, J. B., *An Analysis of Radio Interferometry with the Block 0 System*, JPL Publication 81-49, December 15, 1981.

2. Bartee, T. C., Lebow, I. L., and Reed, I. S., *Theory and Design of Digital Machines*, Fourth Lincoln Laboratory Publication, McGraw-Hill Company, 1962.

3. Mead, C. and Conway, L., *Introduction to VLSI System*, Addison-Wesley Publishing Company, 1980.

4. Ousterhout, J., "Editing VLSI Circuits with Caesar," *Documentation of 1982 VLSI Tools*, Computer Science Division, Electrical Engineering and Computer Science, University of California, Berkeley, April 21, 1982.

5. Terman, C., "ESIM, an Event-Driven Switch Level Simulator for NMOS Transistor Circuits," *Documentation of 1982 VLSI Tools*, Computer Science Division, Electrical Engineering and Computer Science, University of California, Berkeley, April 21, 1982.

6. Negal, L. W. and Pederson, D. O., "SPICE Simulation Program with Integrated Circuit Emphasis," Memorandum No. ERL-M382, Electronics Research Laboratory, University of California, Berkeley, April, 1973.

7. Deutsch, L. J. and Lahmeyer, C. R., "A Systolic Architecture for the Correlation and Accumulation of Digital Sequences," *TDA Progress Report, 42-85*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 62–68, January–March 1986.

8. Deutsch, L. J., "An Integrated UNIX-Based CAD System for the Design and Testing of Custom VLSI Chips," *TDA Progress Report, 42-81*, Jet Propulsion Laboratory, Pasadena, Calif., pp. 51–62, January–March 1986.

Table 1. The change of state of a 3-bit binary updown counter and _U_ and _D_ values

D	U	$A_3(t)$	$A_2(t)$	$A_1(t)$	$A_3(t+1)$	$A_2(t+1)$	$A_1(t+1)$
0	1	0	0	0	0	0	1
0	1	0	0	1	0	1	0
0	1	0	1	0	0	1	1
0	1	0	1	1	1	0	0
0	1	1	0	0	1	0	1
0	1	1	0	1	1	1	0
0	1	1	1	1	0	0	0
0	1	1	1	0	1	1	1
1	0	0	0	0	1	1	1
1	0	0	0	1	0	0	0
1	0	0	1	0	0	0	1
1	0	0	1	1	0	1	0
1	0	1	0	0	0	1	1
1	0	1	0	1	1	0	0
1	0	1	1	0	1	0	1
1	0	1	1	1	1	1	0
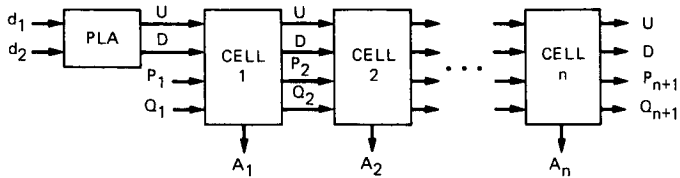
Fig. 1. Pipeline architecture for an *n*-bit binary updown counter

Fig. 3. Logic diagram of a basic cell of an *n*-bit updown counter

d_1	d_2	U	D
0	0	0	0
0	1	0	1
1	0	1	0

Fig. 2. Block diagram and truth table
of a data conversion PLA
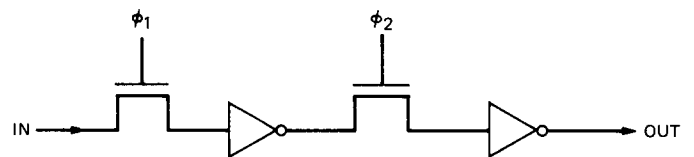
Fig. 4. Logic diagram of a 1-bit shift register
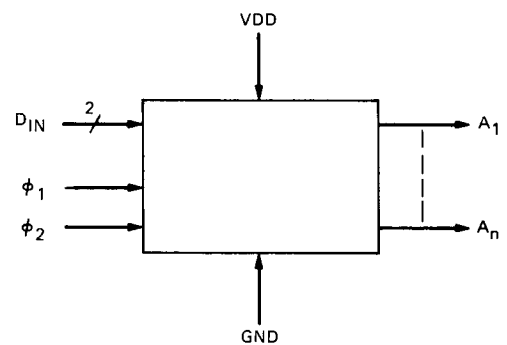
Fig. 5. Pin assignment diagram of an *n*-bit
binary updown counter
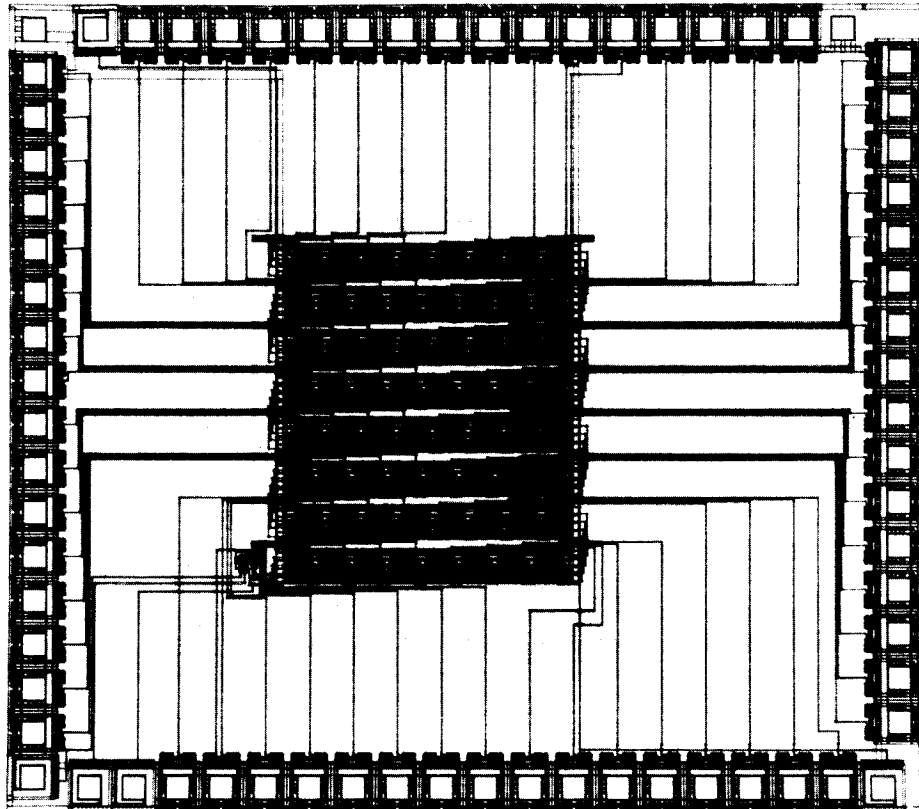
Fig. 6. VLSI layout of a 64-bit updown counter

Filter Distortion Effects on Telemetry Signal-to-Noise Ratio

R. Sadr and W. Hurd

Communications Systems Research Section

The effect of filtering on the Signal-to-Noise Ratio (SNR) of a coherently demodulated band-limited signal is determined in the presence of worst-case amplitude ripple. The problem is formulated mathematically as an optimization problem in the L_2-Hilbert space. The form of the worst-case amplitude ripple is specified, and the degradation in the SNR is derived in a closed form expression. It is shown that when the maximum passband amplitude ripple is 2Δ (peak-to-peak), the SNR is degraded by at most $(1 - \Delta^2)$, even when the ripple is unknown or uncompensated. For example, an SNR loss of less than 0.01 dB due to amplitude ripple can be assured by keeping the amplitude ripple to under 0.42 dB.

I. Introduction

Amplitude ripple is inherent in most physically realizable filters. We seek to determine the system performance degradation resulting from amplitude ripple. The Signal-To-Noise Ratio (SNR) is used as the performance criterion. Our results provide an easy method to determine the worst-case loss due to amplitude ripple. To derive the expression for the worst-case loss in the SNR, the class of worst-case amplitude ripple is explicitly found.

We consider filtering a signal $y(t)$ which is composed of a bandlimited signal $s(t)$, with bandwidth $(f_0 - W, f_0 + W)$ Hertz, added to a noise process $\{n(t)\}$, where f_0 denotes the center frequency and W is the half bandwidth of the signal. The passband of the filter covers the same band of frequency as the signal $s(t)$. This filter is shown in Fig. 1. Ideally, the transfer function for the filter is

$$H_0(j\omega) = e^{j\omega\tau}$$

for $|\omega| \in (2\pi(f_0 - W), 2\pi(f_0 + W))$ where τ is a constant group delay. The shape of the filter response at frequencies outside of the signal band is shown to be immaterial.

The amplitude spectrum of the ideal filter, $H_0(j\omega)$, has constant gain and group delay in the passband frequency. This is never really true in practice where the gain of the filter exhibits a bounded ripple in the passband. The amplitude ripple is the deviation of $|H_0(j\omega)|$ from ideal and is denoted by $\Delta(j\omega)$. Characteristics of both an ideal and a nonideal bandpass filter are depicted in Fig. 1. The transfer function of the nonideal filter in the passband is represented as

$$H_0(j\omega) = (1 + \Delta(j\omega))\, e^{j(\phi(\omega)+\omega\tau)} , \qquad (1)$$

for $|\omega| \in (2\pi(f_0 - W), 2\pi(f_0 + W))$ where $\phi(j\omega)$ represents the deviation in phase from constant group delay, τ.

A critical issue in the specification of the filter $h(t)$, for the design engineer, is to determine the impact of $\Delta(j\omega)$, and

$\phi(j\omega)$ on the SNR. The degradation of SNR for phase deviation $\phi(j\omega)$ has been studied previously. J. Jones (Ref. 1) in 1972 analyzed the filter distortion effects of the phase non-linearity for BPSK and QPSK, and has shown that when phase deviation $\phi(j\omega)$ is bounded by ϕ_{max} in absolute value, the SNR is degraded by at most a factor of $\cos^2\phi_{max}$.

In this article, it is shown that if the amplitude ripple is bounded by Δ (i.e., $|\Delta(j\omega)| \leqslant \Delta$), the SNR in the presence of the worst case amplitude ripple waveform is degraded by at most $(1 - \Delta^2)$. This result holds even when the amplitude ripple is unknown, or known but not compensated.

To summarize the outline of the rest of this article, Section II describes the system under study and the underlying assumptions for which the SNR figure is analyzed. In Section III, the closed form SNR expression is derived for a coherently demodulated signal which is filtered by the non-ideal transfer function characteristic defined in Eq. (1). In Section IV, the class and properties of the worst-case amplitude ripple are specified. Finally, in Section V, we make some concluding remarks based on our results.

II. Formulation

We consider a received waveform containing signal and noise, that is, $y(t) = s(t) + n(t)$. The signal amplitude spectrum $S(j\omega)$ is band-limited to $|f| \epsilon [f_0 + W, f_0 - W]$, and its waveform is completely known during each $t \epsilon [0,T]$. The noise process $\{n(t)\}$ in our analysis is assumed to be an Additive White Gaussian Noise (AWGN) process with single-sided spectral density N_0 W/Hz. The results generalize to the case where the noise is not white. The only restriction is that $\{n(t)\}$ be a wide-sense stationary process. This implies that it has zero mean and autocorrelation function $R_n(\tau) = E[n(t) n(t + \tau)]$, where $E[\cdot]$ denotes the expectation operator.

The optimal receiver for the observed signal $y(t)$, which maximizes the SNR, is a matched filter (Ref. 2). This solution is expressed in the form of the Fredholm integral equation of the first kind. There are known methods to solve this integral equation explicitly to find the optimal matched filter solution $h_{MA}(t)$.

For an AWGN channel the matched filter solution is $h_{MA}(t) = s(T - t)$ or, equivalently, in the frequency domain it is $H_{MA}(j\omega) = S^*(j\omega) e^{-j\omega T}$. (Throughout the article, superscript * denotes complex conjugate while a midline * denotes convolution).

For the case in which the noise is only wide sense stationary (not necessarily AWGN) with spectrum $S_n(j\omega)$, the matched filter solution may be expressed under certain assumptions (Ref. 2) as $H_{MA}(j\omega) = S^*(j\omega) e^{-j\omega T}/S_n(j\omega)$. This transfer function is recognized as the matched filter transfer function for the white noise case divided by the actual power density of the noise. Therefore, it is possible to generalize our result for wide sense stationary noise processes by simply using a matched filter which is matched to both the noise and the known signal. Thus, with no loss of generality, in our subsequent analysis we assume that the noise is white and Gaussian.

In digital communication systems, the signal $s(t)$ is modulated at the transmitter to a Radio Frequency (RF) by multiplying $s(t)$ by the carrier signal $\cos(\omega_0 t)$, where $2\pi \omega_0$ is the carrier frequency. At the receiver (Fig. 2), the observed signal is filtered and then demodulated by multiplying the observed signal by $2 \cos(\omega_0 t)$. This signal is passed through a zonal low-pass filter to filter out the double frequency terms produced by the multiplication operation. The output of the low-pass filter is then fed to the matched filter.

The sampled output of the matched filter each T s is denoted by M_i.

The SNR is defined as the ratio of the square of the expected value to the variance of the random variable M_i. In the following section, a closed form expression for the SNR is derived. This expression is formulated in the form of a functional. We minimize this functional over the ensemble of all possible amplitude waveform ripples in the passband of the filter $h_{BF}(t)$, as shown in Fig. 1.

III. Signal-to-Noise Ratio Expression

Since the filtering processes are linear, we can consider the system response to signal and noise separately. We need to determine the mean value of the signal and the variance of the noise, both at the matched filter output. We denote the response of each stage of the system to the signal by $e_i(t)$, as shown in Fig. 2. The signal is represented by amplitude spectrum throughout the following analysis. Throughout this article, the square of a complex function is meant to be the magnitude square of that function.

Neglecting the noise response, the amplitude spectrum of $y(t)$, $Y(j\omega)$, is expressed as

$$Y(j\omega) = S(j(\omega - \omega_0)) + S(j(\omega + \omega_0)) \qquad (2)$$

We denote the bandpass filter $H'(j\omega)$ as

$$H'(j\omega) = H_{BF}(j(\omega - \omega_0)) + H_{BF}^*(j(-\omega + \omega_0))$$

where $H_{BF}(\cdot)$ is the transfer function of a complex low-pass filter. The spectrum of $e_1(t)$, the output of the bandpass filter under study (Fig. 2), is

$$E_1(j\omega) = H'(j\omega) Y(j\omega) \qquad (3)$$

The demodulated waveform $e_2(t)$ has the spectrum containing the sum and difference frequencies,

$$E_2(j\omega) = E_1(j(\omega - \omega_0)) + E_1(j(\omega + \omega_0))$$

$$= H'(j(\omega - \omega_0)) Y(j(\omega - \omega_0))$$

$$+ H'(j(\omega + \omega_0)) Y(j(\omega + \omega_0)) \qquad (4)$$

The demodulator is followed by an ideal low-pass filter, which filters out the double frequency terms, and the resulting output spectrum $E_3(j\omega)$ is $E_3(j\omega) = E_2(j\omega) H_{LP}(j\omega)$, where

$$H_{LP}(j\omega) = \begin{cases} 1, & |\omega| \leqslant 2\pi(f_0 - W) \\ 0, & \text{otherwise} \end{cases}$$

Thus, the output of the low-pass filter from Eq. (4) is

$$E_3(j\omega) = (H_{BF}^*(-j\omega) + H_{BF}(j\omega)) S(j\omega) \qquad (5)$$

Note that if the low-pass filter is not ideal, its deviation from ideal should be included in the filter under study.

Neglecting deviation in $H_{MA}(j\omega)$ from ideal, the matched filter that maximizes the SNR has the transfer function (Ref. 2) $h_{MA}(t) = s(T - t)$, or equivalently

$$H_{MA}(j\omega) = S^*(j\omega) e^{-j\omega T} \qquad (6)$$

Let for simplicity $H(j\omega)$ denote $H_{BF}^*(-j\omega) + H_{BF}(j\omega)$. Then the matched filter output $e_4(t)$ can be represented as

$$E_4(j\omega) = E_3(j\omega) S^*(j\omega) e^{-j\omega T}$$

and substituting $E_3(j\omega)$ using Eq. (5), $e_4(t)$ may be expressed as

$$e_4(t) = \frac{1}{2\pi} \int_I H(j\omega) |S(j\omega)|^2 e^{-j\omega(T-t)} d\omega \qquad (7)$$

where $I = [-W, W]$. The output of the matched filter is sampled at the end of every time interval T. Thus at $t = T$ we have

$$e_4(T) = \frac{1}{2\pi} \int_I H(j\omega) |S(j\omega)|^2 d\omega$$

The system noise response is denoted by $z(t)$. The random process $\{M_i\}$ (taking values in $\mathbb{R}^1$) is the sum of the filtered signal plus the filtered noise component. Hence, we can write

$$M_i = e_4(T) + z(T) \qquad (8)$$

Taking expectation of Eq. (8), and noting that the noise is assumed to be zero mean, we get

$$E[M_i] = \frac{1}{2\pi} \int_I H(j\omega) |S(j\omega)|^2 d\omega \qquad (9)$$

To compute $\text{Var}[M_i]$, let $z(t) = n(t) * x(t)$. Note that $z(t)$ is a filtered white noise process, which is filtered by the filter under study and the matched filter. The cascaded filter is denoted by $x(t)$.

Let $X(j\omega) = S(j\omega) H(j\omega) e^{-j\omega T}$, thus we have

$$\text{Var}[M_i] = \text{Var}[z(t)] \Big|_{t=T} \qquad (10)$$

and

$$z(t) = n(t) * x(t)$$
$$E[n(t) n(t + \tau)] = N_2 \delta(\tau) \qquad (11)$$

From Eq. (11), the variance of $z(t)$ can be expressed as

$$\text{Var}[z(t)] = \frac{N_0}{4\pi} \int_I |X(j\omega)|^2 d\omega \qquad (12)$$

Combining Eqs. (12) and (10) and evaluating these expressions at $t = T$, results in

$$\text{Var}[M_i] = \frac{N_0}{4\pi} \int_I |H(j\omega) S(j\omega)|^2 d\omega \qquad (13)$$

Therefore since $\text{SNR} = (E[M_i])^2 / \text{Var}[M_i]$ we have

$$\text{SNR} = \frac{2}{N_0} \frac{\left(\int_I H(j\omega) S^2(j\omega) d\omega \right)^2}{\int_I |H(j\omega) S(j\omega)|^2 d\omega} \qquad (14)$$

The filter transfer function is $H(j\omega) = H_{BF}(j\omega) + H_{BF}^*(-j\omega)$ and furthermore $H(j\omega)$ is

$$H(j\omega) = (1 + \Delta(j\omega)) e^{j\omega\tau} \quad \forall \; |\omega| \leqslant 2\pi W$$

We assume that the group delay is negligible (i.e., $e^{j\omega\tau} \approx 1$), or it is compensated in the matched filter, hence, by substituting $H(j\omega)$ in the SNR expression (14), the SNR is

$$\text{SNR} = \frac{2}{N_0} \frac{\left| \int_I (1 + \Delta(j\omega)) S^2(j\omega) \, d\omega \right|^2}{\int_I (1 + \Delta(j\omega))^2 S^2(j\omega) \, d\omega} \quad (15)$$

IV. Worst Case Amplitude Ripple

With no loss of generality assume $I = [0,1]$ and let

$$\xi = \int_0^\xi S^2(j\omega) \, d\omega$$

Using Eq. (15) we can formulate the minimization problem

$$F(g) = \inf_{g(\xi) \in L_{2[0,1]}} \frac{\left(\int_0^1 (1 + g(\xi)) \, d\xi \right)^2}{\int_0^1 (1 + g(\xi))^2 \, d\xi} \quad (16)$$

subject to the constraint

$$|g(\xi)| < \Delta < 1 \quad (16a)$$

The integrals are understood in the Lebesgue sense.

In the appendix we prove the following theorem. In the proof, the class of the functions for which the minimum occurs is explicitly exhibited.

Theorem 1. There is a continuum of measurable step functions which minimize Eq. (16), and at the minimum $F_{MIN}(\tilde{g}) = 1 - \Delta^2$, and $\tilde{g}(\xi)$ is

$$\tilde{g}(\xi) = \begin{cases} \Delta, & \text{for } 0 < \xi \leqslant \dfrac{1 - \Delta}{2} \\ -\Delta, & \text{for } \dfrac{1 - \Delta}{2} < \xi < 1 \end{cases}$$

Using the result of Theorem 1, we can state the following corollary.

Corollary 1. Define interval I_1, as any subset of the interval I such that

$$\int_{I_1} S^2(j\omega) \, d\omega = \frac{1 - \Delta}{2}$$

Then Eq. (15) is minimized by

$$\Delta(j\omega) = \begin{cases} \Delta, & \text{for } \omega \epsilon \, I_1 \\ -\Delta, & \text{for } \omega \notin I_1 \end{cases}$$

Corollary 2. The minimum SNR for Eq. (15) is achieved by $\Delta(j\omega)$ of the form specified by corollary (1), and furthermore, the total SNR for the worst case ripple distortion is $(1 - \Delta^2) \, \text{SNR}_{\text{ideal}}$.

Corollaries (1) and (2) are direct consequences of Theorem 1. The point a, in Fig. 3, indicates the point at which the integrated power of the signal $s(t)$ is $(1 - \Delta)/2$.

In general, the SNR is minimized when the ripple is $+\Delta$ for frequencies containing $(1 - \Delta)/2$ of the energy of $s(t)$, and $-\Delta$ for frequencies containing $(1 + \Delta)/2$ of the energy of $s(t)$.

The shape of the amplitude ripple is not unique, and it is the whole continuum of step functions which satisfies the condition stated in corollary (1). To construct another amplitude ripple waveform which satisfies the conditions of corollary (1), one can take the waveform of Fig. 3 and move a segment from $[0,a]$ to $(a,1]$, and move an equal energy segment from $(a,1]$ to $[0,a]$. Conceptually, this method may be thought of as juggling equal energy line segments from each interval. This method results in obtaining a new step function from the basic function of Fig. 3 which satisfies the statement of corollary (1).

V. Discussion and Conclusion

We have shown that the worst-case loss in SNR due to amplitude ripple is $(1 - \Delta^2)$. This result can be used to specify filters for communication receivers, such as the Advanced Receiver for the NASA's Deep Space Network.

To express the maximum allowable ripple, for a given loss (L_{dB}) in dB, we expand the expression for dB of ripple and

keep the first terms. Thus for small ripple and loss, the ripple loss in dB may be expressed as

$$\text{Ripple}_{dB} \ (L_{dB}) = 4.168 \sqrt{L_{dB}}$$

Consider a simple example: If the system can tolerate a loss of 0.01 dB due to amplitude ripple, the maximum allowable ripple in dB would be 0.4168 dB. The loss is much less than the ripple and decreases as the square of the ripple.

From a mathematical standpoint, we solved the minimization problem stated in Eq. (16). And we showed that the optimal solution lies on the boundary, and it is a continuum of measurable step functions in the interval [0,1]. This is explicitly exhibited in the appendix.

Acknowledgment

References

1. Jones, J. J., "Filtered Distortion and Intersymbol Interference Effects on PSK Signals," *IEEE Trans. Comm. Tech., COM-19*, No. 2, April 1971: 121-132.

2. Van Trees, H. L., *Detection, Estimation, and Modulation Theory*. New York: John Wiley & Sons, Part I, 1968; Part II, 1971.
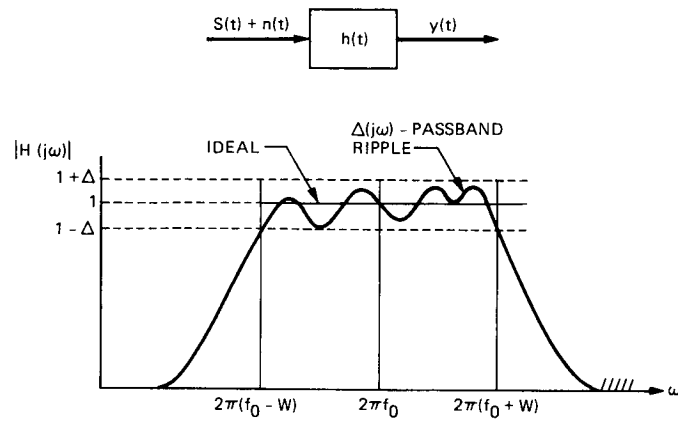
Fig. 1. Filtering
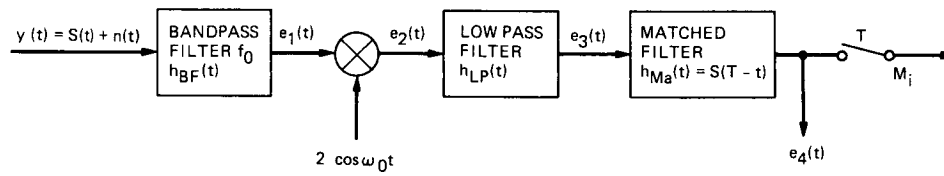
Fig. 2. Communication model
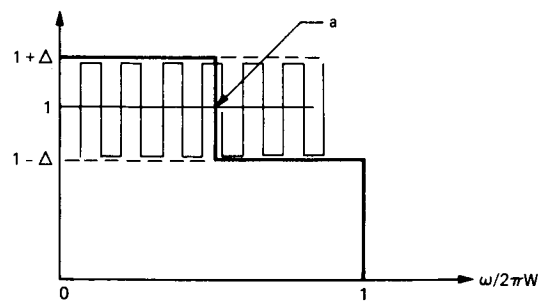
Fig. 3. Ripple waveform

Appendix

Let $0 < \Delta < 1$ and

$$C_\Delta = \left\{ f : [0,1] \to \mathbb{R} \mid f \text{ measurable} \right.$$

$$\left. \text{and } |f(z) - 1| < \Delta \text{ for almost all } u \right\}$$

We equip C_Δ with the induced metric from $L^1(I)$ $(I = [0,1])$ so that C_Δ becomes a complete metric space (we identify functions equal almost everywhere). Consider the continuous functional F on C_Δ defined by

$$F(f) = \frac{\left(\displaystyle\int_0^1 f(x)\,dx \right)^2}{\displaystyle\int_0^1 f^2(x)\,dx}$$

In the appendix we investigate the existence and nature of minima of F. Let

$$C_n = \left\{ (a_0, \ldots, a_n; c_1, \ldots, c_n) \mid \right.$$

$$\left. 0 = a_0 \leqslant a_1 \leqslant \ldots \leqslant a_n = 1, \, c_j \in [1 - \Delta, 1 + \Delta] \right\}$$

Then C_n is naturally identified with a compact subset of $[0,1]^{n-1} \times [1 - \Delta, 1 + \Delta]^n$. To every $\gamma \in C_n$ we assign the function $f_\gamma \in C_\gamma$ which takes value c_i on the open interval (a_{i-1}, a_i) provided it is nonempty. (Note that f_γ is undefined on a finite set, and can be assigned arbitrary values there.) The mapping

$$\gamma \longrightarrow F(f_\gamma)$$

is a continuous function on C_n and hence achieves a minimum on C_n.

Lemma. Let $\gamma = (a_0, \ldots, a_n; c_1, \ldots, c_n)$ and assume $F(f_\gamma)$ is a minimum on C_n. Then $c_i = 1 \pm \Delta$ whenever $a_{i-1} < a_i$.

Proof. Let C'_{n+2} be the compact subset of C_{n+2} consisting of $(a_0, \ldots, a_{n+2}; c_1, \ldots, c_{n+2})$ such that

$$c_1 = 1 + \Delta \quad \text{and} \quad c_{n+2} = 1 - \Delta$$

We embed C_n in C'_{n+2} by

$$(a_0, \ldots, a_n; c_1, \ldots, c_n)$$

$$\longrightarrow (a_0, a_0, \ldots, a_{n-1}, a_n, a_n; 1 + \Delta, c_1, \ldots, c_n, 1 - \Delta)$$

and prove the assertion for C'_{n+2}. Let

$$M_\gamma = \frac{\displaystyle\int_0^1 f_\gamma^2(x)\,dx}{\displaystyle\int_0^1 f_\gamma(x)\,dx}$$

Assume f_γ assumes value $c \neq 1 \pm \Delta$ on a nonempty open interval (a_{i-1}, a_i). Either $c \geqslant M_\gamma$ or $c < M_\gamma$ and assume for definiteness that the former possibility occurs. We may assume $c_2 = c \geqslant M_\gamma$. In fact it is easy to see that there is $\gamma' \in C_{n+2}$ such that $M_{\gamma'} = M_\gamma$, $F(f_\gamma) = F(f_{\gamma'})$ and $c'_2 = c$ where $\gamma' = (a'_0, a'_1, \ldots, a'_{n+2}; c'_1, \ldots, c'_{n+2})$. Let $0 \leqslant \delta \leqslant |a_2 - a_1|$ and define $\Theta \in C'_{n+2}$ by

$$\Theta = (a_0, a_1 + \delta, a_2, \ldots, a_{n+2}; 1 + \Delta, c_2, \ldots, c_{n+1}, 1 - \Delta)$$

Then by a simple computation

$$F(f_\Theta) = \frac{A + \alpha}{B + \beta}$$

where

$$A = \left(\int_0^1 f_\gamma(x)\,dx \right)^2$$

$$B = \int_0^1 f_\gamma^2(x)\,dx$$

and

$$\frac{\alpha}{\beta} = \frac{\displaystyle\int_0^1 f_\gamma(x)\,dx + \frac{1}{2}(1 + \Delta - c)\,\delta}{c + \frac{1}{2}(1 + \Delta - c)}$$

Since $c \geqslant M_\gamma$, by taking $\delta > 0$ sufficiently small we can ensure $\alpha/\beta < A/B$ and consequently $F(f_\Theta) < F(f_\gamma)$. This proves the lemma.

Let $\{f_n\}$ be a sequence in C_Δ such that

$$\lim_{n \uparrow \infty} F(f_n) = \inf_{f \epsilon C_\Delta} F(f)$$

Since

$$\bigcup_n \{f_\gamma \mid \gamma \epsilon C_n\}$$

is dense in C_Δ there is sequence $\{\gamma_k\}$ with $\gamma_k \epsilon C_{n_k}$ such that

$$F(f_{\gamma_k}) < F(f_k) + \frac{1}{k}$$

In view of the lemma we may assume f_{γ_k} takes only values $1 \pm \Delta$ and hence there is $\Theta_k \epsilon C_2$ such that $F(f_{\gamma_k}) = F(f_{\Theta_k})$.

Therefore $\inf F(f)$ is achieved for some $F(f_\Theta)$ with $\Theta = (0, a, 1; 1 + \Delta, 1 - \Delta)$. It is a simple exercise to show that $a = (1 - \Delta)/2$ and $F(f_\Theta) = 1 - \Delta^2$.

Finally we note that for any partition $I = E_1 \cup E_2$ with $\text{meas}(E_1) = (1 - \Delta)/2$, $\text{meas}(E_2) = (1 + \Delta)/2$ the function

$$f(x) = \begin{cases} 1 + \Delta, & \text{if } x \epsilon E_1 \\ 1 - \Delta, & \text{if } x \epsilon E_2 \end{cases}$$

is also a minimum for F. The argument above also shows that all minima of F are of this form.

Fixed Lag Smoothers for Carrier Phase and Frequency Tracking

R. Kumar[1] and W. J. Hurd

Communications Systems Research Section

The article presents the application of fixed lag smoothing algorithms to the problem of estimation of the phase and frequency of a sinusoidal carrier received in the presence of process noise and additive observation noise. A suboptimal structure consists of a phase-locked loop (PLL) followed by a post loop correction to the phase and frequency estimates. When the PLL is operating under a high signal-to-noise ratio, the phase detector is approximately linear, and the smoother equations then correspond to the optimal linear equations for an equivalent linear signal model. The performance of such a smoother can be predicted by linear filtering theory. However, if the PLL is operating near the threshold region of the signal to noise ratio, the phase detector cannot be assumed to be linear. Then the actual performance of the smoother can significantly differ from that predicted by linear theory. In the article we present both the theoretical and simulated performance of such smoothers derived on the basis of various models for the phase and frequency processes.

I. Introduction

The derivation of optimum receivers through modern estimation techniques has been proposed by various researchers (see Refs. 1–12 and their references). In Refs. 4 and 5 optimum zero lag receivers have been derived on the basis of linear Kalman filtering theory (Ref. 6) for linear measurement schemes. The nonlinear measurement situations which are of interest here have been studied in Ref. 7, wherein, on the basis of nonlinear filters of Ref. 8, suboptimal nonlinear zero lag receivers have been derived for the demodulation of angle modulated signals. In Ref. 9 the techniques of Refs. 7 and 8 have been extended to design suboptimum fixed lag smoothers for phase estimation. The solution of the optimum nonlinear filtering/smoothing problem is, of course, intractable. Whether derived from linear or nonlinear theory, the smoother structure consists of a phase-locked loop (PLL) followed by a post loop correction to the phase and frequency estimates.

In this article we study the application of linear and nonlinear smoothers to the phase and frequency estimation of a sinusoid. We show that for this case, the suboptimum nonlinear smoother derived from Refs. 7-9 is not substantially different from the optimum linear smoother equations for an appropriate linearized measurement model. In addition, simu-

---

[1] Also with the faculty of the Electrical Engineering Department at California State University, Long Beach.

lations show that the difference in performance of the non-linear and linear smoother is not significant, even when the phase detector is operating highly nonlinearly. Even though the linear and nonlinear systems perform similarly, linear theory is inadequate to predict the performance when the phase detector is highly nonlinear.

When compared to the case of linear phase detector, the performance of the smoother for the case of nonlinear phase detector can be substantially different. The difference can be much more pronounced when process noise is present, compared to the case when only observation noise is present. The simulation examples indicate that in the absence of process noise, although there is a significant performance degradation due to nonlinearity (about 1 dB when operating in the loop SNR of about 5 dB), there is no threshold observed in the smoothing error covariance in this region. In contrast to this, there is a pronounced threshold in the smoother performance when the process noise is present and the inverse of filter phase error variance is below 7.5 dB.

We also evaluate the smoother performance when the process noise is reduced in magnitude. That is, the smoother/filter solutions are based on a relatively high process noise, but in the simulation the actual variance of the process noise used is lower or zero. This is of interest because, in many practical applications, the process noise statistics are not precisely known, and are therefore deliberately over-estimated.

II. Signal Model and Smoother Equations

In this section, we present a suboptimal nonlinear smoother, and show that it is very similar to the linear solution. We then present an implementation of the linear solution. We consider the problem of estimating the phase process $\theta(k)$ from the sampled version of the received carrier signal $y(k)$, i.e.,

$$y(k) = A\sqrt{2}\sin(\omega_c t_k + \theta(k)) + \bar{v}(k) \tag{1}$$

where t_k is the kth sampling time, ω_c is the known carrier frequency and $\bar{v}(k)$ the observation noise is the sampled version of a narrow band zero mean white Gaussian noise process $v(t)$. Furthermore, the phase process $\theta(k)$ is modeled as

$$\theta(k) = \beta\ell'x(k), \qquad \ell' = [1\ 0\ \ldots\ 0] \tag{2}$$

$$x(k+1) = \Phi x(k) + w(k)$$

In Eq. (2), β is the phase constant, $x(k)$ is the state vector of dimension n, Φ is an $(n \times n)$ matrix and $w(k)$ is zero mean white Gaussian noise process independent of $\{\bar{v}(k)\}$. Thus

$$E[\bar{v}(k)] = 0 \ , \quad E[w(k)] = 0$$

$$E[\bar{v}^2(k)] = R \ ; \quad E[w(k)w^T(k)] = Q \ ; \quad E[\bar{v}(k)w(j)] = 0$$

As shown in Refs. 9 and 10, when $2\omega_c t$ and higher order harmonic terms are ignored, the smoother equations reduce to the following:

$$
\left.
\begin{aligned}
\hat{x}_0(k+1) &= \Phi\hat{x}_0(k/k) + K_0(k+1)\eta(k+1) \\[6pt]
\hat{x}_i(k+1/k+1) &= \hat{x}_{i-1}(k/k) + K_i(k+1)\eta(k+1) \\[6pt]
\eta(k+1) &= \sqrt{2}y(k+1)\cos(\hat{\Theta}(k+1)) \\[6pt]
\hat{\Theta}(k+1) &= \omega_c t_{k+1} + \beta\ell'\hat{x}(k+1/k) \\[6pt]
x_i(k) &\triangleq x(k-i), \qquad i = 0, \ldots, L
\end{aligned}
\right\} \tag{3}
$$

where $\hat{x}_i(k/k)$ represents the filtered estimate of $x_i(k)$ or the smoothed estimate $\hat{x}(k-i/k)$. The gain vectors K_i and the cross covariance matrices $P_{i0}(k/j) \triangleq E\{\tilde{x}(k-i/j)\ \tilde{x}^T(k/j)\}$ with $\tilde{x}(k/j) \triangleq x(k) - \hat{x}(k/j)$ are given by

$$K_i(k+1) = A\beta P_{i0}(k+1/k)\ell\,S^{-1}(k+1), \qquad 0 \leqslant i \leqslant L \tag{4}$$

$$P_{i0}(k+1/k+1) = P_{i0}(k+1/k)$$
$$-P_{i0}(k+1/k)(A\beta\ell)(A\beta\ell)'$$
$$\times P_{00}(k+1/k)S^{-1}(k+1) \tag{5a}$$

$$P_{i0}(k+1/k) = P_{i-1,0}(k/k)\Phi', \qquad 0 < i \leqslant L \tag{5b}$$

$$S^{-1}(k+1) = A^{-2}\left(p_\phi - \frac{1}{2}p_\phi^2\right)^{-1}$$
$$\times \left\{1 - \left[\frac{\tilde{R}(k+1) + p_\phi^2}{\tilde{R}(k+1) + 2p_\phi}\right]^{1/2}\right\}$$

$$p_\phi = \beta^2 P_{00}{}^{1,1}, \qquad \tilde{R}(k) = R(k)/A^2 \tag{6}$$

The smoother error covariance matrix $P_{ii}(k+1/k+1)$ is

$$P_{ii}(k+1/k+1) = P_{ii}(k+1/k) - P_{i0}(k+1/k)(A \beta \ell)$$

$$\times S^{-1}(k+1)(A \beta \ell)' P_{i0}'(k+1/k)$$

$$P_{ii}(k+1/k) = P_{i-1,i-1}(k/k) \tag{7}$$

A. Rapprochement With Linear Theory

Representing the bandpass additive noise $\bar{v}(k)$ in terms of its baseband quadrature components $\bar{v}_i(k)$ and $\bar{v}_q(k)$, ignoring the $2\omega_c$ term, then for small estimation error $\tilde{x}(k+1/k)$,

$$\eta(k+1) = \beta A \ell' \tilde{x}(k+1/k) + \frac{1}{\sqrt{2}} \bar{v}_i(k+1) \tag{8}$$

The $\eta(k+1)$ given by Eq. (8) above is precisely the one-step ahead prediction error (innovation) for the following linear model

$$y(k+1) = \beta A \ell' x(k+1) + \frac{1}{\sqrt{2}} \bar{v}_i(k+1) \tag{9}$$

It is easily verified that Eqs. (3–7) reduce to the linear optimal smoother equations for the model (Eqs. [2,9]), under the assumption of small p_ϕ.

B. Smoother Implementation

If the various gains are replaced by their respective steady state values, the smoother consists of a digital phase-locked loop followed by a post-loop correction to the filtered estimates. As shown in Ref. 10 this post-loop correction can be equivalently implemented by a finite impulse response (FIR) filter whose output $\epsilon(k+1)$ is related to its input $\eta(k+1)$, as in Fig. 1, by

$$\epsilon(k+1) = \eta(k+1) + \gamma^{-1} \eta(k) + \ldots + \gamma^{-(L-1)} \eta(k+2-L) \tag{10}$$

The scalar γ in Eq. (10) can be expressed in terms of steady state filter error covariance matrix, etc., as shown in Ref. 10.

III. Linear Filter/Smoother: Derivation of Transfer Functions and Performance Expressions

In Ref. 10, three specific cases of model (2) are considered. These correspond to the dimension n of the state vector x in Eq. (2) equal to 1, 2, and 3. The resulting filter/smoother configurations are termed first order, second order and third order respectively. By replacing various gains and matrices by

their steady-state values in Eqs. (3–7), these difference equations are replaced by algebraic equations and may be solved explicitly for the steady state values of the filter error covariance matrix P_F, the prediction error covariance P_p, smoother error covariance P_S, etc. Substitutions of these expressions in Eq. (4) and the linearized version of Eq. (6) results in the steady-state expressions for the filter and smoother gains. Finally, from Eq. (3) the filter and smoother transfer functions and various other transfer functions of interest are derived. These expressions are very useful in evaluating the error performance of the filter/smoother when the design value of the process noise covariance matrix is different than its actual value. In such cases, the expressions derived for the filter/smoother covariance matrices do not reflect the actual performance. The various error variances are instead evaluated using frequency domain techniques from the derived transfer functions. One may refer to Ref. 10 for the details of such derivations.

The above derivations are based on the assumption of linear phase detector. The performance predicted on the basis of these expressions is compared with simulations in the next section. As would be observed there, under the assumption of linear phase detector, the simulation results are in close conformity with those predicted from theory.

IV. Simulation Results

In the following the simulation results obtained for the second-order case are presented in some detail. We discuss the performance of the optimal linear filter and smoother both with linear phase detectors first. Then we evaluate the smoother performance versus delay and lastly discuss the smoother performance with the nonlinear phase detector. To be concrete we use the following often used model for the Q matrix

$$Q = \begin{bmatrix} T^2/3 & T/2 \\ T/2 & 1 \end{bmatrix} \sigma_a^2 T^2 \tag{11}$$

One advantage of using the above Q is that the performance of the filter/smoother is then a function of only three parameters viz. σ_a, $\sigma_{\bar{v}}$, and T, where $\sigma_{\bar{v}}^2$ denotes the noise variance of $v_i(k)/\sqrt{2}$ in the baseband model, Eq. (9). We present the smoother/filter performance in terms of the phase estimation error. One may refer to Ref. 10 for the corresponding results for the frequency estimation error.

First we present the phase tracking performance from both analysis and simulations for the case when the phase

detector nonlinearity is ignored. Both the optimal filter and smoother performance are analyzed in the following.

1. Optimal filter and smoother with large delay.

Here we present the performance of the optimal filter and smoother (assuming linear signal model) so as to relate these parameters to the two-sided normalized loop noise bandwidth $2B_p$, a commonly used parameter in the design of phase locked loops. We consider the smoother with large delay. In terms of the closed-loop transfer function matrix $G_F(z)$ (Ref. 10), the parameter $2B_p$ is given by,

$$2B_p = \frac{1}{2\pi j} \int_\Gamma G_{F,p}(z)\, G_{F,p}(z^{-1})\, \frac{dz}{z} \qquad (12)$$

where $G_{F,p}$ represents the first component of the transfer function matrix $G_F(z)$, and Γ is some appropriate contour of integration. Figure 2 plots B_p as calculated from Eq. (12) as a function of $\left(\sigma_a^2/\sigma_v^2\right)$ for three different values of the sampling period T viz. 0.01, 0.1 and 1 s. From these graphs it is readily seen that approximately,

$$2B_p \cong 0.67 \left(\sigma_a^2/\sigma_v^2\right)^{0.11} T, \qquad T = 1 \text{ s}, \sigma_a^2/\sigma_v^2 \leqslant 10$$

$$2B_p \cong 1.01 \left(\sigma_a^2/\sigma_v^2\right)^{0.226} T, \qquad T = 0.1 \text{ s}$$

$$2B_p \cong 1.05 \left(\sigma_a^2/\sigma_v^2\right)^{0.25} T, \qquad T = 0.01 \text{ s}$$

$$(13)$$

The last relation may be taken to be the asymptotic relation for $(2B_p/T)$ as $T \to 0$. Figure 2 also includes the normalized loop noise bandwidth of the smoother as calculated from Eq. (12) with $G_{F,p}$, replaced by the first component of the smoother transfer function matrix. Denoted by $2B_{p,S}$ this normalized bandwidth is given by

$$2B_{p,S} \cong 0.28 \left(\sigma_a^2/\sigma_v^2\right)^{0.234} T, \qquad T = 0.1 \text{ s} \qquad (14a)$$

Comparison with the filter bandwidth of Eq. (13) yields

$$B_p/B_{p,S} \cong 3.6 \left(\sigma_a^2/\sigma_v^2\right)^{-0.008}, \qquad T = 0.1 \text{ s} \qquad (14b)$$

This indicates that the improvement achievable by using second order smoothing compared to a second order filter is approximately a factor of 3 · 6 or 5.5 dB.

The real two-sided noise bandwidth of the filter $2B_{Lp} = 2B_L/T$ and is equal to the normalized phase error variance $P_F(1,1)/\left(\sigma_v^2 T\right)$ when the actual process variance $\sigma_{a,S}^2 = 0$. Similarly

$$2B_{Lp,S} = 2B_{p,S}/T = P_S(1,1)/\left(\sigma_v^2 T\right)$$

In Fig. 3 we plot the normalized phase error variance for both the filter and smoother as obtained from the recursive solutions of Eqs. (4–7). From the figure approximate expressions for these terms may be written as,

$$P_F(1,1)/\sigma_v^2 T \cong 0.75 \left(\sigma_a^2/\sigma_v^2\right)^{0.08}, \quad T = 1 \text{ s}, \ \sigma_a^2/\sigma_v^2 \leqslant 10$$

$$\cong 1.32 \left(\sigma_a^2/\sigma_v^2\right)^{0.22}, \quad T = 0.1 \text{ s}$$

$$\cong 1.4 \left(\sigma_a^2/\sigma_v^2\right)^{0.25}, \quad T = 0.01 \text{ s} \qquad (15a)$$

$$P_S(1,1)/(\sigma_v^2 T) = 0.365 \ \sigma_a^2/\sigma_v^2 \Big.^{0.237} \qquad (15b)$$

Comparing Eqs. (15) and (13,14) one observes that provided an optimum filter or smoother is used, the maximum degradation of the phase error variance is only about 1.34 (1.25 dB) and this is almost independent of the variance σ_a^2.

2. Optimal smoother performance with linear phase detector.

In Fig. 4 is plotted the smoother performance evaluated from simulations as a function of the smoother delay and the ratio $\left(\sigma_a^2/\sigma_v^2\right)$ used in the smoother design, assuming linear phase detector. The dotted curves in the figure plot the two-sided normalized loop noise bandwidth as computed from Eq. (12). As may be inferred from the figure, the two measures of performance are equal within the limits of statistical errors. The minimum phase error variance (corresponding to $L = \infty$) varies over a range of about 0.3 to 1.4 for $\left(\sigma_a^2/\sigma_v^2\right)$ between 1 to 100. It is also apparent from the figure the fact that the number of delays required has an inverse relation to $\left(\sigma_a^2/\sigma_v^2\right)$ to achieve asymptotic smoother performance. In Fig. 5 is plotted the real loop noise bandwidth $2B_{Lp,S}$ as a function of normalized smoother delay (LT/τ_F), where τ_F is the time constant of the optimal filter. As is clear from the figure, the normalized value of delays required to achieve asymptotic smoother performance does not depend significantly upon $\left(\sigma_a^2/\sigma_v^2\right)$.

Figures 6 and 7 plot the results similar to those of Figs. 4 and 5 respectively, when the actual process noise variance

equals its design value, i.e., $\sigma_{a,S}^2 = \sigma_a^2$. A comparison of these two sets of figures shows that the phase error variance can be at most 1.35 times more than for the case of $\sigma_{a,S}^2 = 0$. For intermediate values of the noise variance, $0 < \sigma_{a,S}^2 \leqslant \sigma_a^2$, the ratio would be smaller.

In Fig. 8, a comparison of smoother phase error variance is made for three different sampling periods T equal to $0.01, 0.1$ and 1 s respectively. As is evident from the figure, whereas the optimum filter performance is dependent upon the sampling period, the asymptotic smoother performance depends only marginally on T. Thus the smoother in the most part compensates for any loss of optimality due to finite sampling period. This means that smaller T can be used with smoothing than with filtering only.

3. Smoother performance with nonlinear phase detector.

Figures 9 and 10 present the simulation results for the case of $\left(\sigma_a^2 = \sigma_{\bar{v}}^2\right)$ corresponding to a two-sided noise bandwidth of the filter equal to 1 Hz. For this filter design and in the absence of the process noise $\left(\sigma_{a,S}^2 = 0\right)$, the effect of nonlinearity is to degrade the normalized phase error variance $P_S(1,1)/\sigma_{\bar{v}}^2$ by at most a factor of 1.32 for $\sigma_{\bar{v}}^2 \leqslant 2.2$ (corresponding to the filter rms phase error of $27°$ for linear detector and $30°$ for nonlinear phase detector).

For the case of $\sigma_{a,S}^2 = \sigma_a^2$, the normalized phase estimation error variance depends much more strongly on $\sigma_{\bar{v}}^2$. For $\sigma_{\bar{v}}^2 < 1.4$ (corresponding to rms phase error of 28.5 degrees at the phase detector output), the degradation is within a factor of 1.7 (2.3 dB). The degradation can be much higher for larger values of $\sigma_{\bar{v}}^2$.

From these simulations it may also be inferred that for $\sigma_{a,S}^2 = 0$, and for the case of linear phase detector, the smoother provides an improvement of 5.6 dB over the filter. When the phase detector nonlinearity is taken into consideration, then for a range of $10 \log (1/\sigma_\phi^2) \geqslant 6$ dB, with σ_ϕ^2 denoting phase error variance of the PLL with linear phase detector, the

smoother still provides an improvement of at least 5.1 dB over the filter. Note, however, that the filter performance can itself be degraded by as much as 1.5 dB due to phase detector non-linearity. Since these results correspond to a fixed value of B_L, it may be concluded that with a smoother, the receiver can be operated with at least 3.5 dB smaller carrier power to noise spectral density ratio (P_c/N_0) when it is desired to have 0.1 or smaller value for the phase error variance.

For the case of $\sigma_{a,S}^2 = \sigma_a^2$, it is observed that the effects of nonlinearity are more dominant resulting in a threshold behavior in the smoother phase error variance. However, for $10 \log (P_c/N_0 \, B_L) \geqslant 7.5$ dB, the results in terms of smoother performance are close to those for the case of $\sigma_{a,S}^2 = 0$.

V. Conclusions

The article has presented the performance of suboptimal filter and smoother for the phase and frequency estimation of a sinusoidal carrier under the presence of both the process noise and observation noise. The performance predicted on the basis of linear estimation theory is in close conformity with the corresponding results obtained with simulations, when the phase detector is assumed linear. Similar results are applicable when the phase detector nonlinearity is taken into account and the receiver is operating under high SNR conditions. Under these conditions the smoother improves both the phase and frequency estimation error compared to the filter by about 6 dB.

However, as the SNR is reduced, the corresponding improvement is less. Also the reduction is more when the process noise is present than when only the observation noise is present. Overall taking into account the degradation caused by the nonlinearity in the performance of filter, the smoother can permit the receiver operation at about 3.5 dB smaller carrier power to noise spectral density ratio when it is desired to have 0.1 or smaller value of the phase error variance.

Acknowledgment

References

1. D. C. Youla. "The Use of the Method of Maximum Likelihood in Estimating Continuous-Modulated Intelligence Which has been Corrupted by Noise," *IRE Trans. Inform. Theory, PGIT-3*, pp. 90-105, March 1954.

2. H. L. Van Trees, "Analog Communication Over Randomly-Time-Varying Channels," *IEEE Trans. Inform. Theory, IT-12*, pp. 51-63, January 1966.

3. A. J. Viterbi, "On the Minimum Mean Square Error Resulting from Linear Filtering of Stationary Signals in White Noise," *IEEE Trans. Inform. Theory, IT-11*, pp. 594-595, October 1965.

4. H. L. Van Trees, "Application of State-Variable Techniques in Detection Theory," *Proc. IEEE, 58*, pp. 653-669, May 1970.

5. D. L. Snyder, *The State Variable Approach to Analog Communication*. Cambridge, Mass.; MIT Press, 1971.

6. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Problems," *J. Basic Eng., Ser. D, Vol. 83*, pp. 95-108, March 1961.

7. C. N. Kelly and S. C. Gupta, "Discrete-Time Demodulation of Continuous-Time Signals," *IEEE Trans. Inform. Theory, IT-18*, pp. 448-493, July 1972.

8. A. M. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, New York: 1970.

9. S. Prasad and A. K. Mahalanabis, "Finite Lag Receivers for Analog Communication," *IEEE Trans. Comm., COM-23*, pp. 204-211, February 1975.

10. R. Kumar, *Optimum Filters and Smoothers Design for Carrier Phase and Frequency Tracking*, to appear as JPL Publication, Jet Propulsion Laboratory, Pasadena, Calif., 1987.

11. C. A. Pomalaza Raez and W. J. Hurd, "Improved Carrier Tracking by Smoothing Estimates," *IEEE Trans. Aerosp. Electron. Syst., AES-21*, pp. 610-618, September 1985.

12. B. Ekstrand, "Analytical Steady-State Solutions for Kalman Tracking Filter," *IEEE Trans. on Aerosp. Electron. Syst., 19*, pp. 815-819, November 1983.

Fig. 1. An equivalent and simpler implementation of the smoother

Fig. 2. Normalized loop noise bandwidth of the second order filter/smoother vs (σ_a^2/σ_v^2)
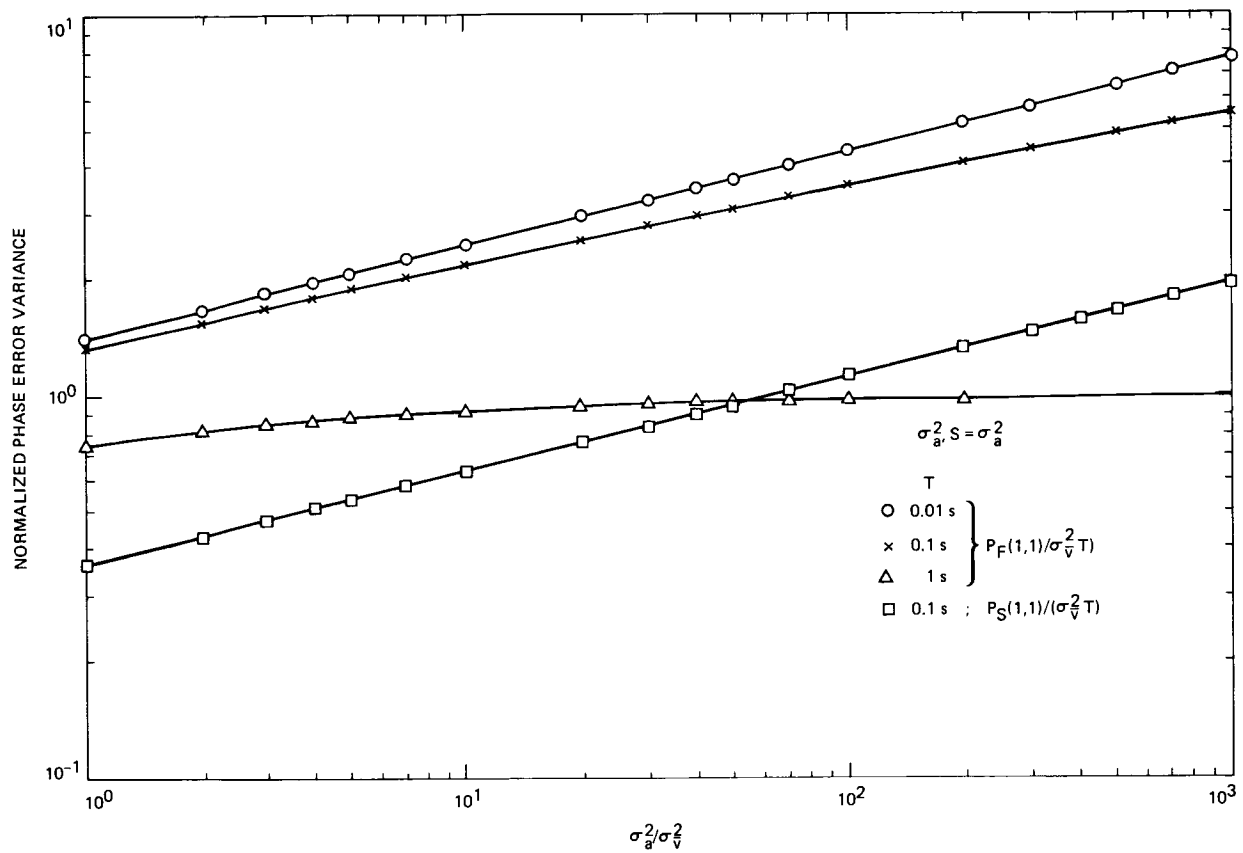
Fig. 3. Normalized phase error variance vs (σ_a^2/σ_v^2) with $\sigma_{a,S}^2 = \sigma_a^2$
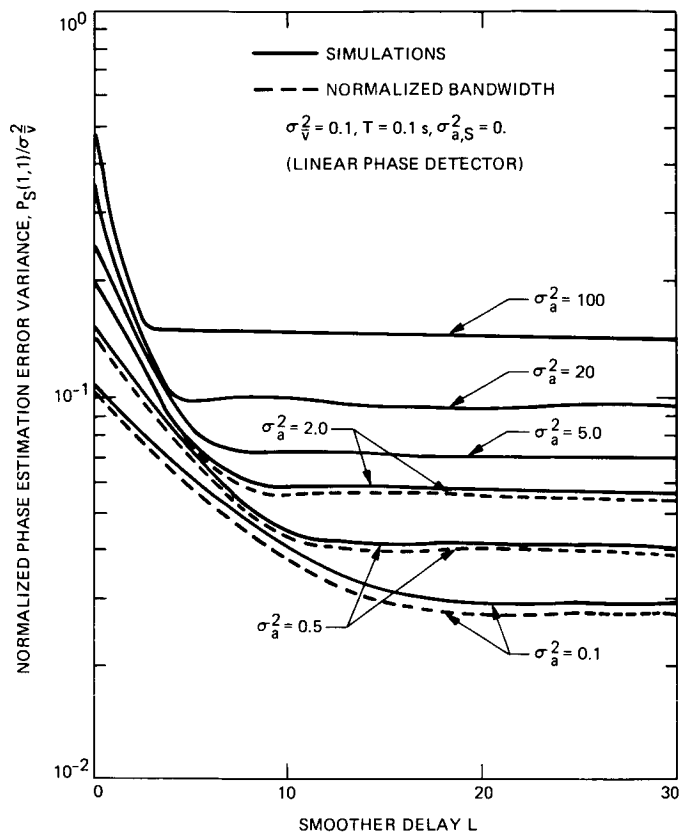
Fig. 4. Smoother estimation error variance vs smoother delay
$(\sigma^2_{a,S} = 0)$
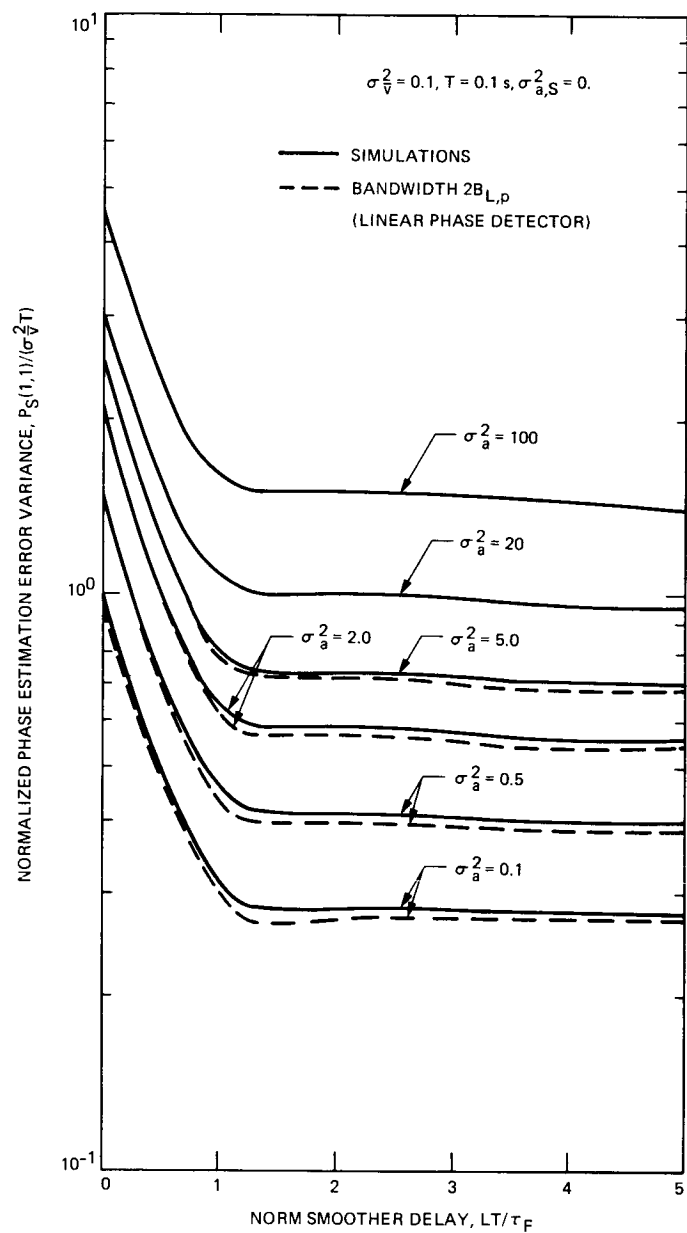
Fig. 5. Normalized smoother estimation error variance vs normalized delay $(\sigma^2_{a,S} = 0)$

Fig. 6. Smoother estimation error variance vs smoother delay $(\sigma_{a,S}^2 = \sigma_a^2)$
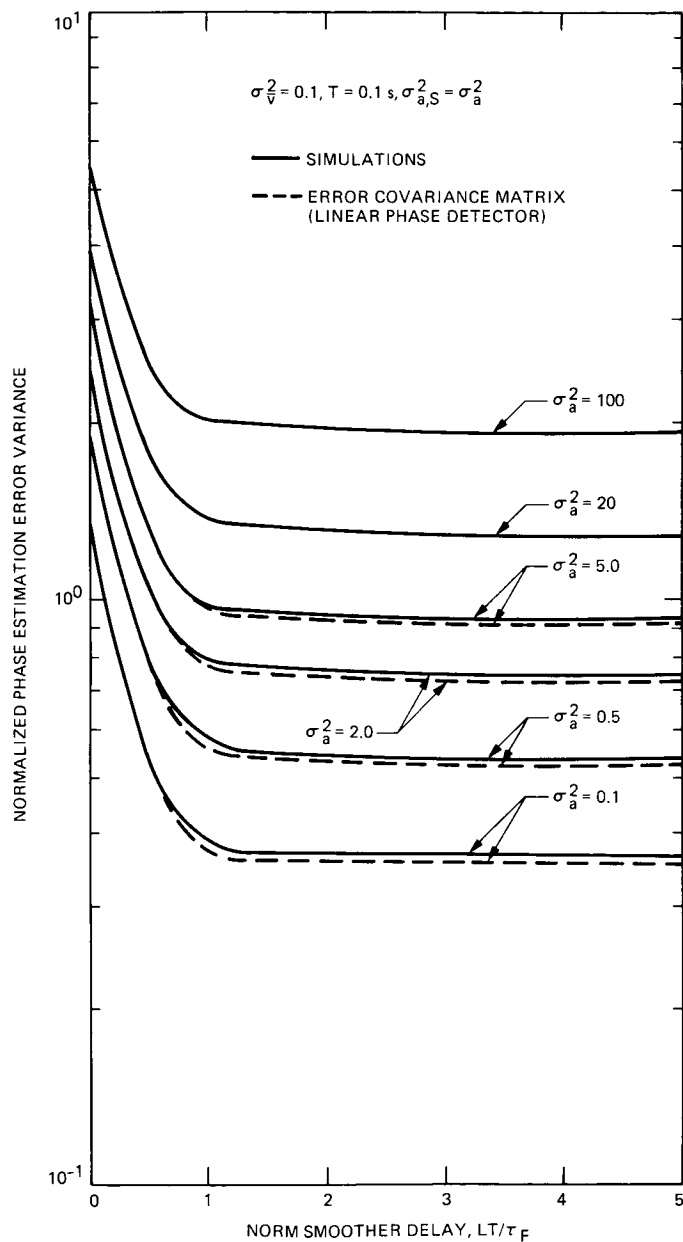
Fig. 7. Normalized smoother estimation error variance vs normalized delay $(\sigma_{a,S}^2 = \sigma_a^2)$
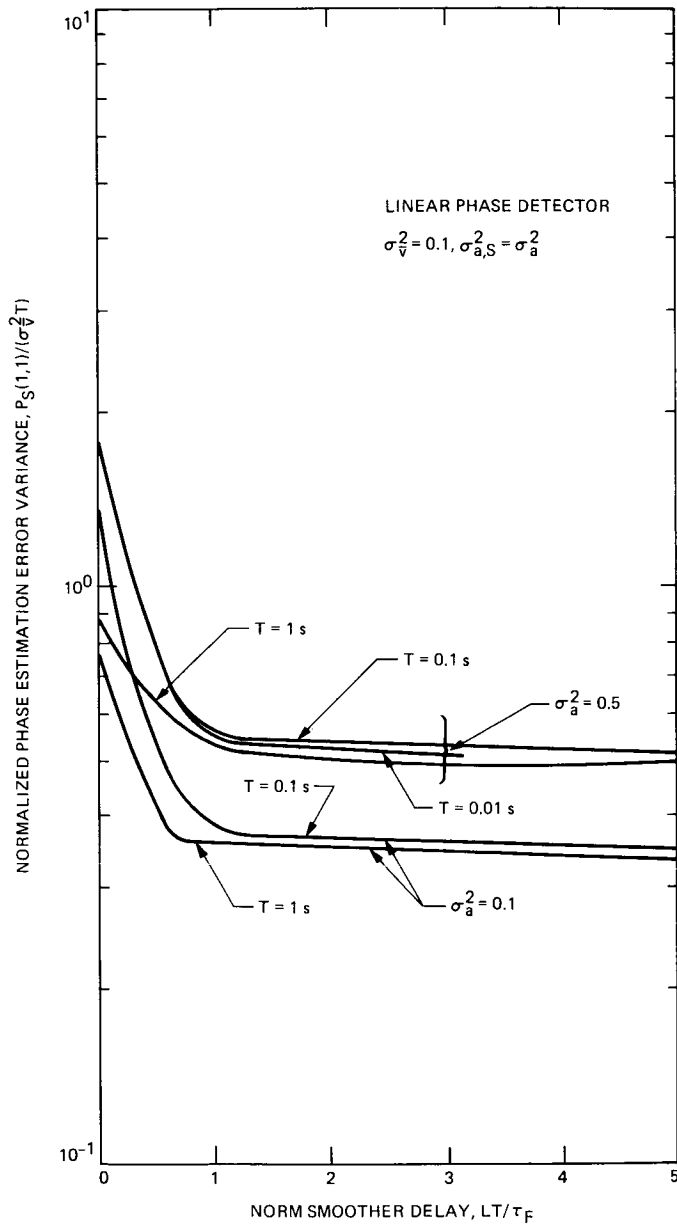
Fig. 8. Smoother performance for different sampling periods
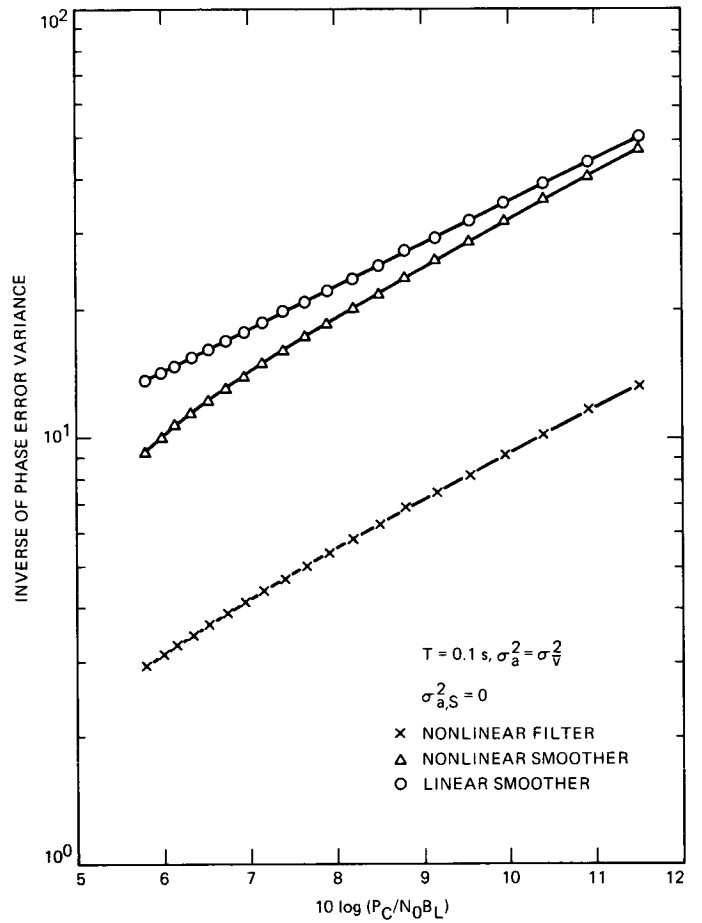
Fig. 9. Comparison of smoother performance with linear and nonlinear phase detectors ($\sigma^2_{a,S} = 0$)
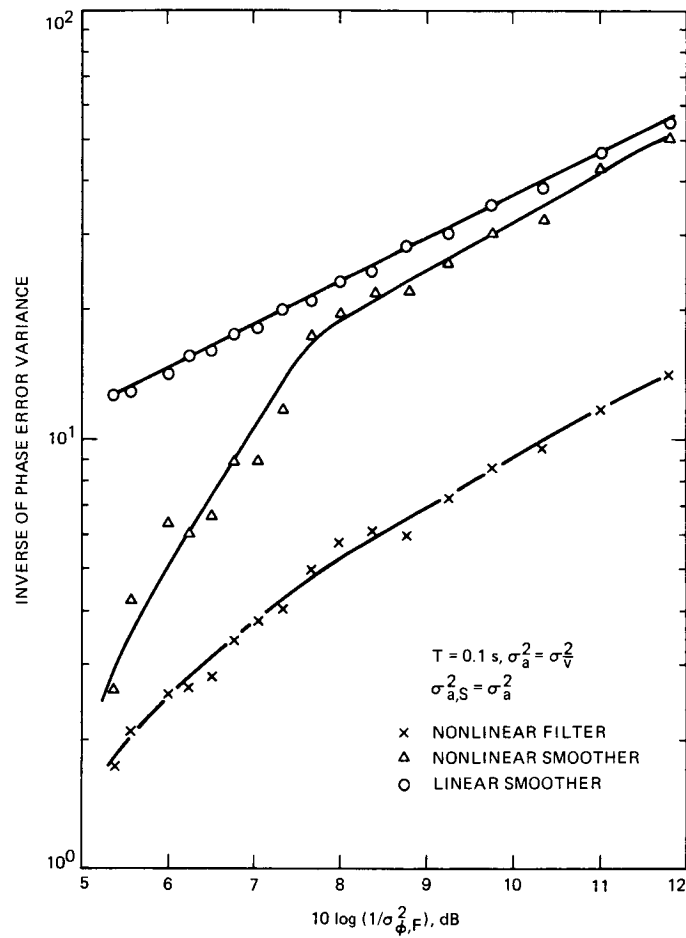
Fig. 10. Comparison of smoother performance with linear and nonlinear phase detectors $(\sigma^2_{a,S} \neq 0)$

An Algorithm to Design Finite Field Multipliers Using a Self-Dual Normal Basis

C. C. Wang

Communications Systems Research Section

Finite field multiplication is central in the implementation of some error-correcting coders. Massey and Omura [4] have presented a revolutionary design for multiplication in a finite field. In their design, a normal basis is utilized to represent the elements of the field. In this article, the concept of using a self-dual normal basis to design the Massey-Omura finite field multiplier is presented. The article first presents an algorithm to locate a self-dual normal basis for $GF(2^m)$ for odd m. Then a method to construct the product function for designing the Massey-Omura multiplier is developed. It is shown that the construction of the product function based on a self-dual basis is simpler than that based on an arbitrary normal basis.

I. Introduction

Finite field multiplication is central in the implementation of some error-correcting coders [1] [2] and authentication devices [3]. There is a need for good multiplication algorithms that can be easily realized. Massey and Omura [4] have developed a new algorithm for multiplication in a Galois field based on a normal basis representation. Using this normal basis, the design of the finite field multiplier is simple and regular [5]. The product components can be obtained by the same logical function operating on the cyclically shifted versions of the components of the multiplicand and multiplier. Hence, designing a Massey-Omura multiplier is essentially designing this product function. An architecture for implementing Massey-Omura multipliers in $GF(2^m)$ was presented in [5]. The normal basis used in the design of [5] is the linearly independent roots of a generating polynomial of $GF(2^m)$. However, it is very difficult to verify the linear independence of the roots of a polynomial. Wah and Wang [6] [7] have shown that if $m + 1$ is a prime and 2 is primitive mod $(m + 1)$, the all-one polynomial of degree m is irreducible and its roots constitute a normal basis. Pei, Wang and Omura [8] have also presented necessary and sufficient conditions for an element to generate a normal basis for the field $GF(2^m)$ for some particular m's. Recently a generalized algorithm to locate a normal basis in any field has been developed [9]. In [9], the concept of dual basis is used to design the product function of the Massey-Omura multiplier.

In this article, a self-dual normal basis is used to design the Massey-Omura multiplier. It is well known [1] that there

exists a self-dual normal basis in $GF(2^m)$ if m is odd. This article will show that the construction of the product function for a self-dual normal basis is simpler than that for an arbitrary normal basis. It also presents an algorithm to locate a self-dual normal basis in $GF(2^m)$ for odd m. Finally, a method to construct the product function is developed.

II. Massey-Omura Finite Field Multiplier

The fundamental concept of Massey-Omura finite field multiplication [4] [5] [9] is based on the utilization of a normal basis of the form $\{\alpha, \alpha^2, \alpha^4, \cdots, \alpha^{2^{m-1}}\}$. Multiplication in the normal basis representation requires the same logic circuitry for any one product component as it does for any other product component. Adjacent product-component circuits differ only in their inputs, which are cyclically shifted versions of one another.

Let $\{\alpha, \alpha^2, \alpha^4, \cdots, \alpha^{2^{m-1}}\}$ be a normal basis for $GF(2^m)$. Any two elements y and z in $GF(2^m)$ can be expressed as

$$y = y_0\alpha + y_1\alpha^2 + y_2\alpha^{2^2} + \cdots + y_{m-1}\alpha^{2^{m-1}}$$

$$= \sum_{i=0}^{m-1} y_i\alpha^{2^i} \tag{1}$$

$$z = z_0\alpha + z_1\alpha^2 + z_2\alpha^{2^2} + \cdots + z_{m-1}\alpha^{2^{m-1}}$$

$$= \sum_{i=0}^{m-1} z_i\alpha^{2^i} \tag{2}$$

Let

$$\omega = y \cdot z$$

$$= \omega_0\alpha + \omega_1\alpha^2 + \omega_2\alpha^{2^2} + \cdots + \omega_{m-1}\alpha^{2^{m-1}}$$

$$= \sum_{k=0}^{m-1} \omega_k\alpha^{2^k} \tag{3}$$

Then, as stated in [4] [5] [9],

$$\omega_{m-1} = f(y_0, y_1, y_2, \cdots, y_{m-1};$$

$$z_0, z_1, z_2, \cdots, z_{m-1})$$

$$\omega_{m-2} = f(y_{m-1}, y_0, y_1, \cdots, y_{m-2};$$

$$z_{m-1}, z_0, z_1, \cdots, z_{m-2})$$

$$\vdots \tag{4}$$

$$\omega_1 = f(y_2, y_3, \cdots, y_{m-1}, y_0, y_1;$$

$$z_2, z_3, \cdots, z_{m-1}, z_0, z_1)$$

$$\omega_0 = f(y_1, y_2, \cdots, y_{m-1}, y_0;$$

$$z_1, z_2, \cdots, z_{m-1}, z_0)$$

where

$$f(a_0, a_1, \cdots, a_{m-1}; b_0, b_1, \cdots, b_{m-1})$$

$$= \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \rho_{ij} a_i b_j \tag{4a}$$

with $\rho_{ij} = 0$ or 1. Therefore, the central problem in designing a Massey-Omura multiplier is to construct the product function f given in (4a). A product function can be constructed in such a way that the coefficient ρ_{ij} of $a_i b_j$ in (4a) is

$$\rho_{ij} = Tr\left(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \gamma^{2^{m-1}}\right) \tag{5}$$

where $Tr(x)$ denotes the trace value of the element x in $GF(2^m)$ and $\{\gamma, \gamma^2, \gamma^{2^2}, \cdots, \gamma^{2^{m-1}}\}$ is the dual basis to the basis $\{\alpha, \alpha^2, \alpha^{2^2}, \cdots, \alpha^{2^{m-1}}\}$ [9].

A simple and equivalent way to represent the product function f is by means of a Boolean matrix

$$\bar{\Omega} = \left[\rho_{ij}\right]_{i,j=0}^{m-1} \tag{6}$$

where the i-j entry ρ_{ij} of $\bar{\Omega}$ is the coefficient of $a_i b_j$ given in (5).

III. Properties of Boolean Matrix Generated by a Self-Dual Normal Basis

A self-dual basis is a basis whose dual basis is itself. It is known [1] that, if m is odd, $GF(2^m)$ has a self-dual normal

$C \cdot 2$

basis. Let $\{\alpha, \alpha^2, \alpha^{2^2}, \cdots, \alpha^{2^{m-1}}\}$ be a self-dual normal basis, that is, $Tr(\alpha^{2^i} \cdot \alpha^{2^j}) = \delta_{ij}$ where $\delta_{ij} = 0$ for $i \neq j$ and 1 for $i = j$. From (5), the Boolean matrix associated with this self-dual normal basis can be written as

$$\overline{\Omega} = \left[\rho_{ij}^*\right]_{i,j=0}^{m-1} \tag{7}$$

where

$$\rho_{ij}^* = Tr\left(\alpha^{2^i} \cdot \alpha^{2^j} \cdot \alpha^{2^{m-1}}\right) \tag{7a}$$

Three properties of the Boolean matrix $\overline{\Omega}$ have been proved in [9]. They are

Property 1

$\overline{\Omega}$ is symmetric, that is, $\rho_{ij}^* = \rho_{ji}^*$.

Property 2

$$\rho_{ii}^* = \begin{cases} 0 \text{ if } i \neq m-2 \\ 1 \text{ if } i = m-2 \end{cases}$$

Property 3

$$\sum_{i=0}^{m-1} \rho_{ij}^* = \begin{cases} 0, & j \neq m-1 \\ 1, & j = m-1 \end{cases}$$

In addition, there are two more properties for the Boolean matrix generated by a self-dual normal basis.

Property 4

$$\rho_{i,m-1}^* = \rho_{m-1,i}^* = \delta_{i0}$$

Proof:

$$\rho_{i,m-1}^* = \rho_{m-1,i}^*$$

$$= Tr\left(\alpha^{2^i} \cdot \alpha^{2^{m-1}} \cdot \alpha^{2^{m-1}}\right)$$

$$= Tr\left(\alpha^{2^i} \cdot \alpha^{2^m}\right)$$

$$= Tr\left(\alpha^{2^i} \cdot \alpha^{2^0}\right) = \delta_{i0} .$$

Property 5

$$\rho_{ij}^* = \rho_{(m-1+i-j)(m-j-2)}^* = \rho_{(j-i-1)(m-i-2)}^*$$

for $i < j$ and $0 \leqslant i, j < m-1$

Proof:

Since $Tr(\alpha) = Tr(\alpha^2)$,

$$\rho_{ij}^* = Tr\left(\alpha^{2^i} \alpha^{2^j} \alpha^{2^{m-1}}\right)$$

$$= Tr\left(\alpha^{2^{[i+(m-1-i)]}} \cdot \alpha^{2^{[j+(m-1-i)]}}\right.$$

$$\left. \cdot \alpha^{2^{[(m-1)+(m-1-i)]}}\right)$$

$$= Tr\left(\alpha^{2^{m-1}} \cdot \alpha^{2^{(m-1-i+j)}} \cdot \alpha^{2^{(2m-2-i)}}\right)$$

$$= Tr\left(\alpha^{2^{m-1}} \cdot \alpha^{2^{(j-i-1)}} \cdot \alpha^{2^{(m-2-i)}}\right)$$

$$= \rho_{(j-i-1)(m-2-i)}^*$$

Also,

$$\rho_{ij}^* = Tr\left(\alpha^{2^{[i+(m-1-j)]}} \cdot \alpha^{2^{[j+(m-1-j)]}}\right.$$

$$\left. \cdot \alpha^{2^{[(m-1)+(m-1-j)]}}\right)$$

$$= Tr\left(\alpha^{2^{(m-1+i-j)}} \cdot \alpha^{2^{m-1}} \cdot \alpha^{2^{(2m-2-j)}}\right)$$

$$= Tr\left(\alpha^{2^{(m-1+i-j)}} \cdot \alpha^{2^{m-1}} \cdot \alpha^{2^{(m-2-j)}}\right)$$

$$= \rho_{(m-1+i-j)(m-j-2)}^*$$

Property 4 implies that the components of the last column and the last row of $\overline{\Omega}$ are all zeros except the first component as shown in Fig. 1. Property 5 illustrates a triangular symmetric structure as shown in Fig.1. This triangular symmetry is described as follows.

Since $\bar{\Omega}$ is symmetric with respect to the diagonal (Property 1), it is sufficient to discuss only the upper-right triangle of $\bar{\Omega}$. Ignoring the main diagonal and the last column, the upper-right triangular portion of the matrix consists of $[m/3]$ *equilateral* triangles in the sense that the numbers of elements on all of the three sides of each triangle are the same. Here, $[x]$ denotes the greatest integer which is smaller than or equal to x. Let Δ_1 denote the outer-most (largest) triangle, and Δ_i the ith outer-most (ith largest) triangle. The triangular symmetric structure is such that the sequences of the vectors counting clockwise on three sides of the triangle Δ_i are identical. Define this identical sequence by $\underline{v}_i$. As shown in Fig. 1, $\underline{v}_i = (a_i, b_i, c_i, \ldots)$ where $a_i, b_i, c_i, \ldots \in GF(2)$. The dimension of $\underline{v}_i$ is $(m-3i)$. As the structure merges toward the inner-most (smallest) triangle, one of the following three possible patterns will happen.

(i) If $m \equiv 0 \bmod 3$, then

(ii) If $m \equiv 1 \bmod 3$, then
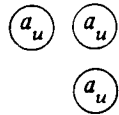
(iii) If $m \equiv 2 \bmod 3$, then
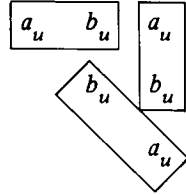
where $a_u, b_u \in GF(2)$, and $u = [m/3]$.

One advantage of using this particular Boolean matrix is that its construction requires fewer trace computations. For $GF(2^m)$, the number of trace computations required to construct this Boolean matrix is

$$\frac{m^2 - 3m + 2}{6}, \quad \text{if } m \neq 0 \bmod 3; \text{ and}$$

$$\frac{m^2 - 3m}{6} + 1, \quad \text{if } m = 0 \bmod 3,$$

which is less than one-third of that required for a Boolean matrix corresponding to an arbitrary normal basis as given in [9].

IV. Locating a Self-Dual Normal Basis in $GF(2^m)$ When m Is Odd

Theorem 26 of Chapter 4 of [1] shows that $GF(2^m)$ has a self-dual normal basis if m is odd. In this section, a method to locate a self-dual normal basis for $GF(2^m)$ when m is odd is presented. Let $\{\underline{\alpha}\} \triangleq \{\alpha, \alpha^2, \alpha^{2^2}, \ldots, \alpha^{2^{m-1}}\}$ be an arbitrary normal basis and $\{\underline{\beta}\} \triangleq \{\beta, \beta^2, \beta^{2^2}, \ldots, \beta^{2^{m-1}}\}$ be a self-dual normal basis in $GF(2^m)$. Then α can be expressed

$$\alpha = b_0\beta + b_1\beta^2 + b_2\beta^{2^2} + \ldots + b_{m-1}\beta^{2^{m-1}} \tag{8}$$

Due to the fact that $\beta^{2^m} = \beta$, one can obtain

$$\begin{bmatrix} \alpha \\ \alpha^2 \\ \alpha^{2^2} \\ \vdots \\ \alpha^{2^{m-1}} \end{bmatrix} = \bar{B} \begin{bmatrix} \beta \\ \beta^2 \\ \beta^{2^2} \\ \vdots \\ \beta^{2^{m-1}} \end{bmatrix} \tag{9}$$

where

$$\bar{B} = \begin{bmatrix} b_0 & b_1 & b_2 & \cdots & b_{m-1} \\ b_{m-1} & b_0 & b_1 & \cdots & b_{m-2} \\ b_{m-2} & b_{m-1} & b_0 & \cdots & b_{m-3} \\ \vdots & \vdots & \vdots & & \vdots \\ b_1 & b_2 & b_3 & \cdots & b_0 \end{bmatrix} \tag{9a}$$

is the transformation matrix from the basis $\{\underline{\beta}\}$ to the basis $\{\underline{\alpha}\}$. Clearly, $\bar{B}$ is invertible. Taking the transpose of (9) results in

$$\left[\alpha, \alpha^2, \alpha^{2^2}, \ldots, \alpha^{2^{m-1}}\right] = \left[\beta, \beta^2, \beta^{2^2}, \ldots, \beta^{2^{m-1}}\right] \bar{B}^T \tag{10}$$

Multiplying (9) by (10), one has

$$
\begin{bmatrix} \alpha \\ \alpha^2 \\ \alpha^{2^2} \\ \vdots \\ \vdots \\ \alpha^{2^{m-1}} \end{bmatrix} \begin{bmatrix} \alpha \ \alpha^2 \ \alpha^{2^2} \ \dots \ \alpha^{2^{m-1}} \end{bmatrix}
$$

$$
= \overline{B} \begin{bmatrix} \beta \\ \beta^2 \\ \beta^{2^2} \\ \vdots \\ \vdots \\ \beta^{2^{m-1}} \end{bmatrix} \begin{bmatrix} \beta \ \beta^2 \ \beta^{2^2} \ \dots \ \beta^{2^{m-1}} \end{bmatrix} \overline{B}^T \qquad (11)
$$

Carrying out the multiplication of the column and row vectors, and then taking the trace function Tr on both sides of (11), it can be shown that

$$
\overline{F}(\alpha) \triangleq \begin{bmatrix} F_{ij} \end{bmatrix}_{i,j=0}^{m-1}
$$

$$
\triangleq \begin{bmatrix} Tr\left(\alpha^{2^i + 2^j}\right) \end{bmatrix}_{i,j=0}^{m-1}
$$

$$
= \overline{B}\,\overline{B}^T \qquad (12)
$$

since $Tr(\beta^{2^i + 2^j}) = \delta_{ij}$.

To locate the self-dual normal basis $\{\beta\}$ from $\{\alpha\}$, $\overline{B}$ of (9a) needs to be solved from (12). Since $\overline{F}_{ij}(\alpha) = F_{\overline{i-1}\ \overline{j-1}}(\alpha)$ where $\overline{j-1} = (j-1) \bmod m$, in (12), it is sufficient to consider only the equality between the first row of $\overline{F}(\alpha)$ and the first row of the product of $\overline{B}\,\overline{B}^T$. Therefore,

$$
Tr(\alpha^2) = b_0^2 + b_1^2 + b_2^2 + \dots + b_{m-1}^2
$$

$$
Tr(\alpha^3) = b_0 b_{m-1} + b_1 b_0 + b_2 b_1 + \dots + b_{m-1} b_{m-2}
$$

$$
\vdots
$$

$$
Tr(\alpha^{2^j+1}) = \sum_{k=0}^{m-1} b_k b_{m-j+k}
$$

$$
\vdots
$$

$$
Tr(\alpha^{2^{m-1}+1}) = b_0 b_1 + b_1 b_2 + b_2 b_3 + \dots + b_{m-1} b_0
$$

$$
(13)
$$

Since $\{\alpha\}$ is a normal basis, $Tr(\alpha^2)$ must be 1. Also, since $b_i \in GF(2), b_i^2 = b_i$. Then the first equation of (13) becomes

$$
1 = b_0 + b_1 + b_2 + \dots + b_{m-1}.
$$

This implies that the set of $\{b_0, b_1, b_2, \dots, b_{m-1}\}$ must have an odd number of 1's.

Applying Lemma 11 of [9], that is, $Tr(\alpha^{2^j+1}) = Tr(\alpha^{2^{m-j}+1})$ for $1 \leqslant j < m/2$, it can be seen that, ignoring the first equation, the first half of the remaining equations in (13) is identical to the second half of the equations in a reverse order. This means that (13) has at most $(m-1)/2 + 1 = (m+1)/2$ linearly independent equations which are

$$
b_0 + b_1 + b_2 + \dots + b_{m-1} = F_{00}(\alpha) = 1
$$

$$
\sum_{k=0}^{m-1} b_k b_{\overline{m-j+k}} = F_{0j}(\alpha)
$$

$$
\text{for } j = 1, 2, 3, \dots, (m-1)/2 \qquad (14)
$$

But, in equation (14), m unknowns $\{b_0, b_1, b_2, \dots, b_{m-1}\}$ need to be found. Therefore, the solution is not unique.

Now, an algorithm to find a solution to (14) is demonstrated by using a simple example of $m = 7$. In this case, (14) becomes

$$
b_0 + b_1 + b_2 + b_3 + b_4 + b_5 + b_6 = 1 = F_{00}
$$

$$
(15a)
$$

$$
b_0 b_1 + b_1 b_2 + b_2 b_3 + b_3 b_4 + b_4 b_5 + b_5 b_6 + b_6 b_0 = F_{01}
$$

$$
(15b)
$$

$$b_0 b_2 + b_1 b_3 + b_2 b_4 + b_3 b_5 + b_4 b_6 + b_5 b_0 + b_6 b_1 = F_{02}$$

(15c)

$$b_0 b_3 + b_1 b_4 + b_2 b_5 + b_3 b_6 + b_4 b_0 + b_5 b_1 + b_6 b_2 = F_{03}$$

(15d)

The purpose of the algorithm is to find a possible solution vector $\underline{b} \triangleq (b_0, b_1, b_2, b_3, b_4, b_5, b_6)$, for a given vector $\underline{t} \triangleq (F_{00}, F_{01}, F_{02}, F_{03})$ under the condition that $F_{00} = 1$. Notice that the left hand sides of equations (15b), (15c) and (15d) are the sums of all possible products $b_i b_{\overline{i+k}}$ $(i = 0, 1, 2, \ldots, 6)$ for $k = 1, 2$, and 3 respectively. Let n_0, n_1, n_2, and n_3 be the numbers of 1's to be added in (15a) (15b), (15c), and (15d), respectively. That is, n_0 is the number of 1's in $\underline{b}$, and n_k $(k = 1, 2, 3)$ is the number of i's such that $b_i b_{\overline{i+k}} = 1$ where $i = 0, 1, 2, 3, 4, 5, 6$. Note that n_0 must be odd. Since $b_i b_{\overline{i+k}} \in GF(2)$, it is clear that, when $F_{0k} = 0, k = 1, 2, 3, n_k$ must be even (considering 0 is even too). On the other hand, when $F_{0k} = 1, n_k$ must be odd.

In this algorithm, b_0 is assumed to be always 1. Since the first element F_{00} of $\underline{t}$ must be 1, eight possible patterns of vector $\underline{t}$ need to be considered.

Case (i): $\underline{t} = (1, 0, 0, 0)$

As $F_{01} = F_{02} = F_{03} = 0, n_1, n_2$ and n_3 must be even. Recognize that the condition that $b_j = 0$ for all $j \neq 0$ can result in $n_1 = n_2 = n_3 = 0$, and, consequently, satisfy the equalities of (15b), (15c) and (15d). Hence, a possible solution $\underline{b}$ is (1, 0, 0, 0, 0, 0, 0).

Case (ii): $\underline{t} = (1, 0, 0, 1)$

As $F_{03} = 1, n_3$ must be odd. Let $n_3 = 1$. From (15d), $b_3 = 1$ can at least satisfy the condition of $n_3 = 1$ (since $b_0 = 1$). Then, a pattern of $\underline{b} = (\underline{1, X, X, 1}, X, X, X)$, where "X" indicates an undecided value, can be temporarily set up. Since n_0 must be odd, there must be at least one, but not an even number of, j's for $j \neq 0$ or 3 such that $b_j = 1$. Let n_0 be the minimum, that is, let there be only one $j(j \neq 0$ or 3) such that $b_j = 1$. Since $F_{01} = F_{02} = 0$, this j must be chosen so that n_1 and n_2 are both even and $n_3 = 1$. In order to satisfy this condition, this particular j must satisfy the condition that $b_{\overline{j-k}} b_j = b_j b_{\overline{j+k}}$ for all k. The only solution for this is that this j is located at the center of a segment which is composed of odd consecutive X's. Hence, $b_5 = 1$, that is, $\underline{b} = (1, X, X, 1, X, \underline{1}, X)$. Now, letting $X = 0$ satisfies the condition that $F_{01} = F_{02} = 0$ and $F_{03} = 1$. Therefore, a solution to (15) is $\underline{b} = (1, 0, 0, 1, 0, 1, 0)$.

Case (iii): $\underline{t} = (1, 0, 1, 0)$

Following the same rules discussed in Case (ii), a solution $\underline{b}$ to (15) can be sequentially decided as

(1) $\underline{b} = (\underline{1, X, 1}, X, X, X, X)$ because $F_{02} = 1$;

(2) $\underline{b} = (1, \underline{1}, 1, X, X, X, X)$ because, as stated in Case (ii), $\underline{b}$ requires an additional "1" to satisfy the condition of n_0 being odd, and, this additional "1" must be located at the center of a segment of odd consecutive X's;

(3) Letting $X = 0, \underline{b} = (1, 1, 1, 0, 0, 0, 0)$.

Case (iv): $\underline{t} = (1, 1, 0, 0)$

Again, a solution $\underline{b}$ can be sequentially decided as

(1) $\underline{b} = (\underline{1, 1}, X, X, X, X, X)$ because $F_{01} = 1$;

(2) $\underline{b} = (1, 1, X, X, \underline{1}, X, X)$ because of the same reasons stated in step (2) of Case (iii);

(3) Letting $X = 0, \underline{b} = (1, 1, 0, 0, 1, 0, 0)$.

Case (v): $\underline{t} = (1, 0, 1, 1)$

As in Case (ii), the first step is to recognize that $F_{03} = 1$. This gives a pattern of $\underline{b} = (\underline{1, X, X, 1}, X, X, X)$. Since the number of elements in $\underline{b}$ is odd, the locations of 1's must divide the present $\underline{b}$ pattern into two segments of consecutive X's. One segment has an even number of X's, while the other has an odd number of X's. The second step is to recognize that $F_{02} = 1$. As in Case (iii), a pattern of $\underline{1, X, 1}$ should exist in $\underline{b}$. Notice that pattern $\underline{1, X, 1}$ has an *odd* number of bits. In order not to affect the equalities given in (15b), (15c) and (15d), this pattern $\underline{1, X, 1}$ must be placed at the center of a segment in $\underline{b}$ which has an *odd* number of consecutive X's. Therefore $\underline{b} = (1, X, X, 1, \underline{1, X, 1})$. Since n_0 must be odd, letting n_0 be the minimum, the third step is to add an additional "1" in $\underline{b}$. Again, following the same argument described in Case (ii), this additional "1" must be placed at the center of a segment with odd consecutive X's, resulting in $\underline{b} = (1, X, X, 1, 1, \underline{1}, 1)$. Finally, letting $X = 0$, a solution $\underline{b} = (1, 0, 0, 1, 1, 1, 1)$ to (15) can be obtained.

Case (vi): $\underline{t} = (1, 1, 0, 1)$

First, since $F_{03} = 1, \underline{b} = (\underline{1, X, X, 1}, X, X, X)$. Next, since $F_{01} = 1$, a pattern of $\underline{1, 1}$ should exist in $\underline{b}$. Similar to what was described in Case (v), since there are an even number of bits in pattern $\underline{1, 1}$, this pattern should be placed at the center of a segment of even consecutive X's in $\underline{b}$. Therefore, $\underline{b}$ becomes $(1, \underline{1, 1}, 1, X, X, X)$. Finally, since n_0 is odd, letting n_0 be the minimum results in $\underline{b} = (1, 1, 1, 1, 0, \underline{1}, 0)$.

Case (vii): $\underline{t} = (1, 1, 1, 0)$

Using the arguments given in Case (v) and Case (vi), $\underline{b}$ can be sequentially decided as

(1) $\underline{b} = (\underline{1, X, 1}, X, X, X, X)$ because $F_{02} = 1$;

(2) $\underline{b} = (1, X, 1, X, \underline{1, 1}, X)$ because $F_{01} = 1$;

(3) $\underline{b} = (1, \underline{1}, 1, X, 1, 1, X)$ because n_0 is odd;

(4) Letting $X = 0, \underline{b} = (1, 1, 1, 0, 1, 1, 0)$, finally.

Case (viii): $\underline{t} = (1, 1, 1, 1)$

Again, $\underline{b}$ can be sequentially decided as

(1) $\underline{b} = (\underline{1, X, X, 1}, X, X, X)$ because $F_{03} = 1$;

(2) $\underline{b} = (1, X, X, 1, \underline{1, X, 1})$ because $F_{02} = 1$;

(3) $\underline{b} = (1, \underline{1, 1}, 1, 1, X, 1)$ because $F_{01} = 1$;

(4) $\underline{b} = (1, 1, 1, 1, 1, \underline{1}, 1)$, because n_0 is odd.

For an arbitrary odd number m, the algorithm of solving $\underline{b} \triangleq (b_0, b_1, b_2, \cdots, b_{m-1})$ to equation (14) for a given $\underline{t} \triangleq (F_{00}, F_{01}, \cdots, F_{0(m-1)/2})$ can be described as follows

(i) Set $b_j = 0$ for $j = 0, 1, \cdots, m - 1$.

Set $k = (m - 1)/2$.

Set ISTART = 0.

(ii) If $F_{0k} = 0$, go to (iii).

Else,

(A) if $b_0 = .0$, set $b_0 = 1, b_k = 1$ and ISTART = k;

(B) if $b_0 = 1$,

(a) if ISTART = odd,

(1) if k = odd,

$b_{(ISTART-k)/2} = 1$ and

$b_{(ISTART+k)/2} = 1$;

(2) if k = even,

$b_{ISTART+(m-ISTART-k)/2} = 1$ and

$b_{m-(m-ISTART-k)/2} = 1$;

(b) if ISTART = even,

(1) if k = odd,

$b_{ISTART+(m-ISTART-k)/2} = 1$ and

$b_{m-(m-ISTART-k)/2} = 1$;

(2) if k = even,

$b_{(ISTART-k)/2} = 1$ and

$b_{(ISTART+k)/2} = 1$.

(iii) Set $k = k - 1$

(iv) If $k \neq 0$, go to (ii).

(v) If ISTART = 0, $b_0 = 1$.

Else, (A) if ISTART = even, $b_{(ISTART+1)/2} = 1$;

(B) if ISTART = odd, $b_{(m+ISTART)/2} = 1$.

(vi) End.

Figure 2 illustrates a flow chart of this algorithm to solve equation (14) when m is odd. It should be pointed out that this algorithm is not the only algorithm for solving $\underline{b}$. However, this algorithm is the optimum in the sense that the number of 1's in $\underline{b}$ is minimum.

Since the matrix $\bar{B}$ in (9) is formed by $\underline{b}$ which depends on only α according to the algorithm described, a self-dual normal basis $(\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}})$ can be constructed from another normal basis $(\alpha, \alpha^2, \alpha^{2^2}, \cdots, \alpha^{2^{m-1}})$ in such a way that

$$
\begin{bmatrix} \beta \\ \beta^2 \\ \beta^{2^2} \\ . \\ . \\ . \\ \beta^{2^{m-1}} \end{bmatrix} = \bar{B}^{-1} \begin{bmatrix} \alpha \\ \alpha^2 \\ \alpha^{2^2} \\ . \\ . \\ . \\ \alpha^{2^{m-1}} \end{bmatrix} \qquad (16)
$$

where $\bar{B}^{-1}$ is the inverse of $\bar{B}$

V. Construction of a Boolean Matrix from a Self-Dual Normal Basis When m is Odd

For an arbitrary element θ in $GF(2^m)$ such that $Tr(\theta) = 1$. $F_{0j} \triangleq Tr(\theta^{2^j+1})$ for $j = 1, 2, \ldots, m - 1$, can be calculated. Following the algorithm described in the last section, a solution $\underline{b} = (b_0, b_1, b_2, \cdots, b_{m-1})$ to the equation (14) can be obtained. Then, a matrix $\bar{B}$ can be constructed in the form of (9a).

Theorem 1

If $\bar{B}$ is invertible, $\{\theta, \theta^2, \theta^{2^2}, \cdots, \theta^{2^{m-1}}\}$ is a normal basis.

Proof:

Since $\bar{B}$ is invertible, $\bar{B}^{-1}$ exists. It can be easily shown that $\bar{B}^{-1}$ must be of form that the row vectors in $\bar{B}^{-1}$ are the

cyclically shifted versions of one another which is the form of $\overline{B}$. Then, $\overline{B}^{-1}$ can be expressed as

$$\overline{B}^{-1} = \begin{bmatrix} b_0' & b_1' & b_2' & \cdots & b_{m-1}' \\ b_{m-1}' & b_0' & b_1' & \cdots & b_{m-2}' \\ b_{m-2}' & b_{m-1}' & b_0' & \cdots & b_{m-3}' \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ b_1' & b_2' & b_3' & \cdots & b_0' \end{bmatrix}$$

Let

$$\beta = \sum_{i=0}^{m-1} b_i' \theta^{2^i}$$

and

$$\underline{\beta} = [\beta, \beta^2, \beta^{2^2}, \beta^{2^3}, \cdots, \beta^{2^{m-1}}].$$

Then

$$\underline{\beta}^T = \overline{B}^{-1} \underline{\theta}^T \qquad (17)$$

where

$$\underline{\theta} = [\theta, \theta^2, \theta^{2^2}, \theta^{2^3}, \cdots, \theta^{2^{m-1}}].$$

Let us first prove that $\{\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}}\}$ is a normal basis. A contradiction proof is used here.

Suppose that $\{\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}}\}$ are linearly dependent, there must exist a non-zero vector $\underline{C} \triangleq [C_0, C_1, C_2, \cdots, C_{m-1}]$ such that

$$\underline{C} \underline{\beta}^T = 0$$

From (17),

$$\underline{C} \overline{B}^{-1} \underline{\theta}^T = 0$$

Since $\underline{\theta}$ is a normal basis, vector $\underline{C}\overline{B}^{-1}$ must be an all-zero vector. This implies that $\{\underline{b}_0', \underline{b}_1', \underline{b}_2', \cdots, \underline{b}_{m-1}'\}$ are linearly dependent where $\underline{b}_k'$ is the kth row vector of $\overline{B}^{-1}$. It contradicts the fact that $\overline{B}^{-1}$ is invertible. Therefore, $\{\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}}\}$ is a normal basis. Since $\underline{\theta}^T = \overline{B} \underline{\beta}^T$ and

$\overline{B}$ is invertible, it is clear that $\{\theta, \theta^2, \theta^{2^2}, \cdots, \theta^{2^{m-1}}\}$ is a normal basis.

Theorem 2

If $\overline{B}$ is invertible, $\{\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}}\}$ which is constructed by (17) is a self-dual normal basis.

Proof:

From Theorem 1, $\{\theta, \theta^2, \theta^{2^2}, \cdots, \theta^{2^{m-1}}\}$ and $\{\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}}\}$ are both normal bases. Following the same procedure of (9)(10) and (11) by replacing α by θ arrives to

$$\overline{F}(\theta) = \overline{B}\, \overline{F}(\beta)\, \overline{B}^T$$

where $\overline{F}(x)$ is an $m \times m$ matrix with entry $F_{ij}(x) = Tr(x^{2^i + 2^j})$, $i, j = 0, 1, \ldots, m-1$. Then,

$$\overline{B}^{-1}\, \overline{F}(\theta)\, (\overline{B}^{-1})^T = \overline{F}(\beta)$$

Since $\underline{b}$ is a solution of $\overline{F}(\theta) = \overline{B}\,\overline{B}^T$, $\overline{B}^{-1}\,\overline{F}(\theta)\,(\overline{B}^{-1})^T = I$. Therefore, $\overline{F}(\beta) = I$, that is $\{\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}}\}$ is a self-dual normal basis.

Now, an algorithm of constructing a Boolean matrix for a self-dual normal basis for $GF(2^m)$ when m is odd can be described as follows.

Starting with an arbitrary element θ in $GF(2^m)$, one first computes $\{F_{00}(\theta), F_{01}(\theta), F_{02}(\theta), \cdots, F_{0,(m-1)/2}(\theta)\}$. Going through the procedure described in the last section, one can obtain a solution $\underline{b} = (b_0, b_1, b_2, \cdots, b_{m-1})$ to equation (14). After forming a matrix $\overline{B}$ as shown in (9a), one checks whether $\overline{B}$ is invertible. If it is, $(\theta, \theta^2, \theta^{2^2}, \cdots, \theta^{2^{m-1}})$ is a normal basis. If it is not, try another θ until the corresponding matrix $\overline{B}$ is invertible. From the normal basis $(\theta, \theta^2, \theta^{2^2}, \cdots, \theta^{2^{m-1}})$, a self-dual normal basis $(\beta, \beta^2, \beta^{2^2}, \cdots, \beta^{2^{m-1}})$ can be formed by (17). Finally applying Property 1, 2, 4 and 5 in section 3, one can compute $\rho_{ij}^* = Tr(\beta^{2^i} \cdot \beta^{2^j} \cdot \beta^{2^{m-1}})$ for $i = 0, 1, \cdots, [m/3] - 1$ and $j = 2i + 1, 2i + 2, \cdots, m - 3 - i$, and then set up the Boolean matrix $\overline{\Omega} = [\rho_{ij}^*]_{i,j=0}^{m-1}$ of structure given in Fig. 1.

Figure 3 illustrates a flow chart of setting up the Boolean matrix. Our initial goal is to determine whether $(\theta, \theta^2, \theta^{2^2}, \cdots, \theta^{2^{m-1}})$ is a normal basis. Theorems 7 and 8 of [9] show two quick check rules to do this before solving the $\underline{b}$. Notice that Theorem 9 of [9] is not applicable here because m is odd. Figures 4(a), (b) and (c) illustrate the Boolean matrices obtained by using this algorithm for $m = 9, 17$ and 31, respectively. [7] also presents the Boolean matrix for $m = 127$. Figure 5 shows the CPU time required on VAX-11/750 to construct the Boolean matrix based on a self-dual normal basis

of $GF(2^m)$. Compared to Fig. 5 of [9], it can be seen that, for large m, the computation time required for a self-dual normal basis is reduced to about 1/3 of that for a regular normal basis. For example, for $m = 127$, it takes only 16 minutes to construct a Boolean matrix versus 40 minutes indicated in Fig. 5 of [9]. This is due to the fact that the number of trace computations required in this algorithm is less than 1/3 of that required in the algorithm stated in [9]. When m is small, the pre-matrix computation which includes the program initial setup and locating the normal basis dominates the computer time. Therefore, Fig. 5 doesn't show significant reduction on computer time when m is small.

In [5], it has been shown that the complexity of the VLSI design of Massey-Omura multipliers depends on the numbers of 1's in Boolean matrix $\bar{\Omega}$. A matrix with fewer 1's is more desirable. Comparing our computer results in this article with those in [9], it is observed that the number of 1's in the Boolean matrix generated by a self-dual normal basis is less than that generated by an arbitrary normal basis.

VI. Conclusion

In this article, it has been shown that the Boolean matrix obtained by a self-dual normal basis maintains a special symmetric structure so that the time required to construct it can be reduced to 1/3 that required for an arbitrary normal basis. To locate a self-dual normal basis in the field $GF(2^m)$ has been a challenging problem. This article has presented an algorithm to locate a self-dual normal basis and then to construct a Boolean matrix when m is odd.

References

1. MacWilliams, F. J., and Sloane, N. J. A., *The Theory of Error-Correcting Codes*, North-Holland Publishing, New York, 1977.

2. Peterson, W. W., and Weldon, Jr., E. J., *Error-Correcting Codes*, MIT Press, Cambridge, 1972.

3. Berkovits, S., Kowalchuk, J., and Schanning, B., "Implementing Public Key Scheme," *IEEE Communications Magazine*, Vol. 17, pp. 2-3, May 1979.

4. Massey, J. L., and Omura, J. K., Patent Application of "Computational Method and Apparatus for Finite Field Arithmetic," submitted in 1981.

5. Wang, C. C., et al., "VLSI Architectures for Computing Multiplications and Inverses in $GF(2^m)$," *IEEE Transactions on Computers*, Vol. C-34, No. 8, pp. 709-717, August 1985.

6. Wah, P. K. S., and Wang, M. Z., "Realization and Application of the Massey-Omura Lock," *Proceedings of International Zurich Seminar*, IEEE Press, pp. 175-182, March 1984.

7. Wang, C. C., *Exponentiation in Finite Fields*, Ph.D. Dissertation, School of Engineering and Applied Science, UCLA, June 1985.

8. Pie, D. Y., Wang, C. C., and Omura, J. K., "Normal Basis of Finite Field $GF(2^m)$," *IEEE Transactions on Information Theory*, Vol. II-32, No. 2, pp. 285-287, March 1986.

9. Wang, C. C., "A Generalized Algorithm to Design Finite Field Normal Basis Multipliers," *TDA Progress Report 42-87*, Jet Propulsion Laboratory, Pasadena, Calif., November 15, 1986, pp. 125-139.

$a_i, b_i, c_i \in GF(2)$
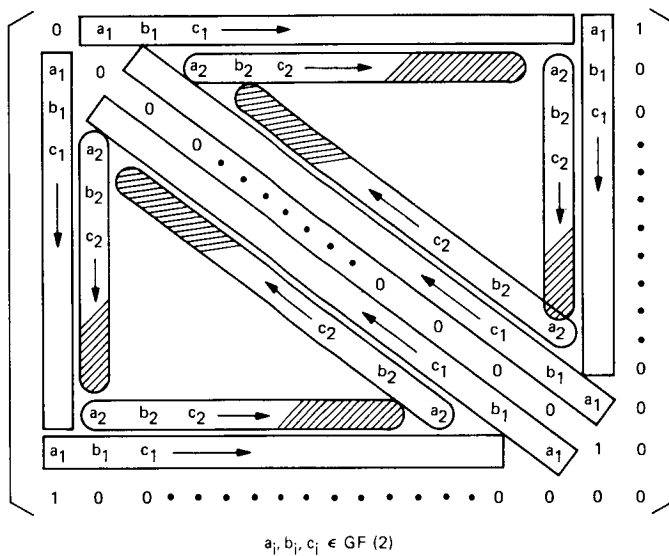
Fig. 1. Structure of a Boolean matrix corresponding to a self-dual normal basis

SET $b_j = 0$ FOR $j = 0, 1, \ldots, m - 1$
SET $k = (m - 1)/2$
SET ISTART = 0

IF $F_{0k} = 0$?

YES / NO

IF ISTART = 0?

YES / NO

$b_0 = 1$
$b_k = 1$
ISTART = k

IF ISTART = ODD?

NO / YES

IF k = ODD?

YES / NO

IF k = ODD?

YES / NO

$b_{(ISTART - k)/2} = 1$
$b_{(ISTART + k)/2} = 1$

$b_{(ISTART + m - k)/2} = 1$
$b_{(ISTART + m + k)/2} = 1$

$k = k - 1$

IF $k = 0$?

NO

IF ISTART = 0?

YES / NO

$b_0 = 1$

IF ISTART = EVEN?

NO / YES

$b_{(m + ISTART)/2} = 1$

$b_{(ISTART + 1)/2} = 1$

END
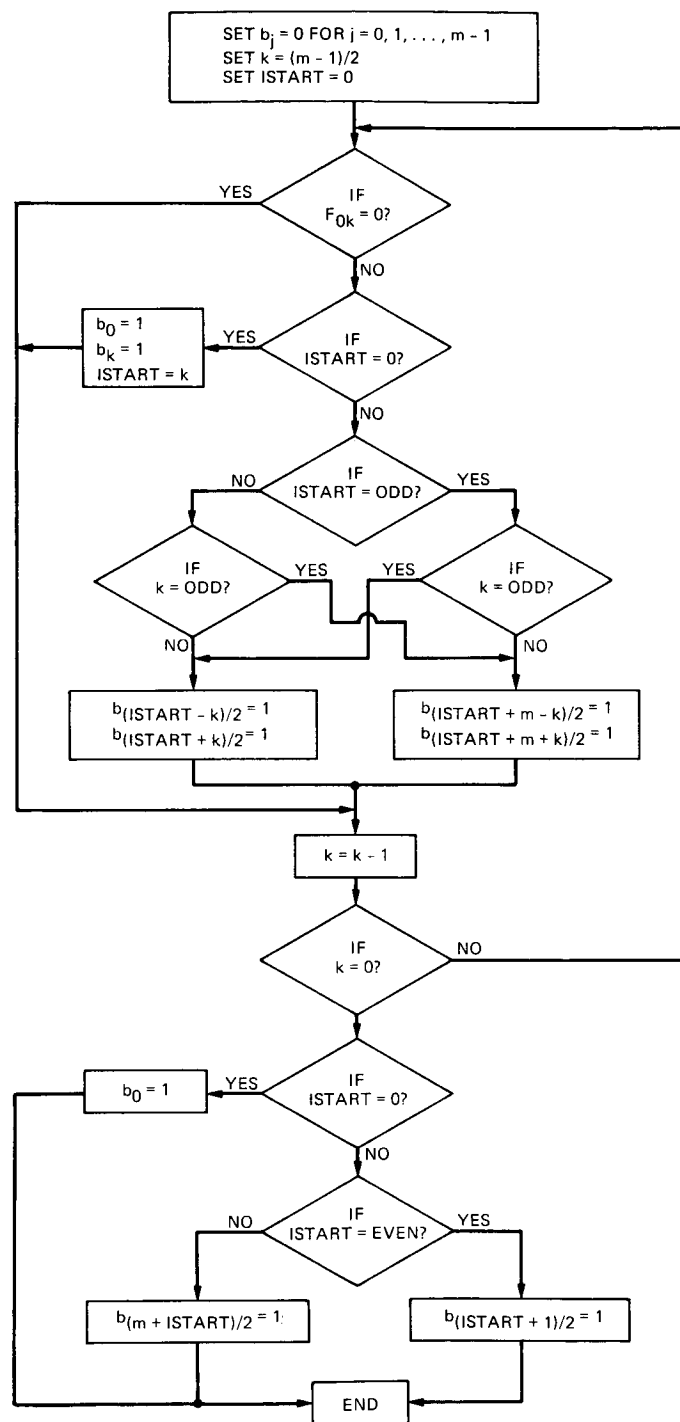
Fig. 2. Flow chart of computing the vector $\underline{b}$
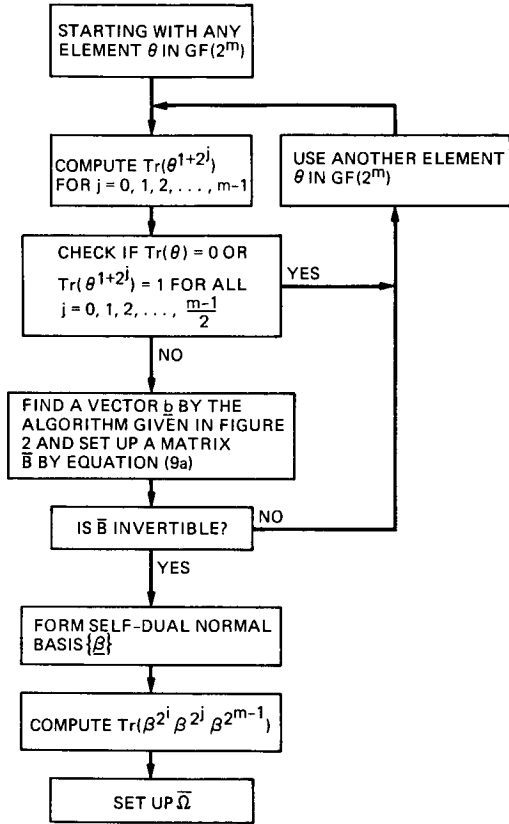
88

(a)
```
  : 0 1 2 3 4 5 6 7 8
 ---------------------
0 : 0 0 0 0 1 1 1 0 1
1 : 0 0 1 0 1 0 0 0 0
2 : 0 1 0 1 0 1 1 0 0
3 : 0 0 1 0 1 1 0 1 0
4 : 1 1 0 1 0 0 0 1 0
5 : 1 0 1 1 0 0 0 1 0
6 : 1 0 1 0 0 0 0 0 0
7 : 0 0 0 1 1 1 0 1 0
8 : 1 0 0 0 0 0 0 0 0
```
NUMBER OF 1's IN BOOLEAN MATRIX = 29

(b)
```
   : 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
   : 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
 -------------------------------------
 0 : 0 1 0 1 0 1 1 0 1 0 1 0 0 0 0 1 1
 1 : 1 0 0 0 1 1 1 1 0 0 1 0 0 0 0 0 0
 2 : 0 0 0 0 0 1 0 0 0 1 0 0 1 1 1 1 0
 3 : 1 0 0 0 0 0 1 0 1 1 0 1 0 0 1 0 0
 4 : 0 1 0 0 0 0 0 0 1 1 1 1 1 0 1 1 0
 5 : 1 1 1 0 0 0 1 1 0 0 1 1 1 0 1 1 0
 6 : 1 1 0 1 0 1 0 0 0 1 1 1 0 1 0 0 0
 7 : 0 1 0 0 0 1 0 0 1 0 0 1 1 0 0 1 0
 8 : 1 0 0 1 1 0 0 1 0 0 1 0 0 0 1 0 0
 9 : 0 0 1 1 1 0 1 0 0 0 0 1 0 1 0 1 0
10 : 1 1 0 0 1 1 1 0 1 1 0 1 1 1 0 0 0
11 : 0 0 0 1 1 1 1 1 0 1 1 0 0 1 0 0 0
12 : 0 0 1 0 1 1 0 1 0 0 1 0 0 1 0 0 0
13 : 0 0 1 0 0 0 1 0 0 1 1 1 1 0 0 0 0
14 : 0 0 1 1 1 1 1 0 0 1 0 0 0 0 0 0 0
15 : 1 0 1 0 1 1 0 1 0 1 0 0 0 0 1 1 0
16 : 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```
NUMBER OF 1's IN BOOLEAN MATRIX = 117

(c)
```
   : 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3
   : 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0
 -----------------------------------------------------------------
 0 : 0 1 1 0 1 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 0 1 0 0 1 1 1 1
 1 : 1 0 1 0 0 1 1 1 0 1 0 1 1 0 1 0 1 1 0 1 1 1 1 1 0 1 0 0 0 1 0
 2 : 1 1 0 0 0 1 1 0 1 0 1 0 0 0 1 0 1 0 0 1 1 0 1 1 0 1 1 1 0 0 0
 3 : 0 0 0 0 0 0 1 1 0 0 0 1 1 1 0 1 0 0 1 1 0 1 1 1 1 1 0 1 1 1 0
 4 : 1 0 0 0 0 1 1 0 0 0 0 1 0 1 1 0 1 0 1 0 1 1 1 0 0 0 0 1 0 0 0
 5 : 0 1 1 0 1 0 0 0 1 1 0 1 1 0 1 0 0 0 1 0 0 0 0 1 1 0 0 1 1 1 0
 6 : 1 1 1 1 1 0 0 1 1 1 1 1 1 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 1
 7 : 1 1 0 1 0 0 0 1 0 1 1 0 1 1 0 0 0 1 0 1 0 0 1 0 0 0 1 1 1 1 0
 8 : 1 0 1 0 0 1 1 1 0 1 1 0 1 1 0 0 1 1 0 0 1 1 1 0 1 0 1 1 0 1 0
 9 : 1 1 0 0 0 1 1 1 1 0 0 1 1 0 0 0 0 1 0 0 0 0 1 0 1 1 0 1 1 0 0
10 : 1 0 1 0 0 0 1 0 1 0 0 0 1 0 1 1 0 1 0 0 0 0 1 0 0 1 0 0 1 1 0
11 : 1 1 1 1 1 0 1 1 0 1 0 0 0 0 0 1 0 1 1 1 1 0 1 0 1 0 0 1 0 0 0
12 : 1 1 0 1 1 1 1 1 1 1 1 0 0 0 1 1 0 0 1 0 0 0 1 0 1 1 1 0 1 1 0
13 : 1 0 0 1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0
14 : 0 1 0 0 1 1 0 0 0 0 1 0 1 0 0 0 0 1 1 1 0 0 0 0 0 1 1 1 1 0 0
15 : 0 0 1 1 1 0 0 0 0 0 1 1 1 1 0 0 0 1 0 0 1 1 0 0 0 0 1 0 1 0 0
16 : 0 1 0 0 0 0 1 1 1 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 1 0 0 0 0 0 0
17 : 0 1 1 0 1 0 0 0 1 0 1 1 1 0 1 1 1 0 1 1 0 1 0 1 1 1 1 1 1 0 0
18 : 0 0 0 1 0 1 1 1 0 1 0 1 0 0 1 0 0 1 0 1 1 1 1 0 1 1 1 0 1 0 0
19 : 0 1 0 1 1 0 1 0 0 0 0 1 0 0 1 0 0 1 1 0 1 0 1 0 0 0 1 0 1 0 0
20 : 0 1 1 0 0 0 0 0 1 0 0 0 0 1 0 1 1 0 1 1 0 1 1 0 0 0 1 1 1 1 0
21 : 1 1 0 1 1 0 0 1 0 0 1 0 0 1 1 1 0 1 0 1 0 1 0 1 0 0 1 0 0 1 1
22 : 1 1 0 1 1 0 0 0 1 0 1 0 0 1 0 0 0 1 1 1 1 1 0 1 1 0 1 0 0 1 0
23 : 1 1 1 1 1 1 1 0 0 1 0 1 1 0 0 0 1 1 0 0 0 0 1 0 1 1 1 1 1 0 0
24 : 0 0 1 1 0 1 1 0 1 0 0 0 1 0 0 0 0 1 1 0 0 1 1 1 0 0 1 1 0 1 0
25 : 1 1 0 0 0 0 0 0 1 0 1 1 0 1 0 1 0 1 1 1 0 0 0 0 1 0 0 1 0 0 0
26 : 0 0 1 1 0 0 0 0 1 1 0 1 0 0 1 0 0 1 1 1 1 1 0 1 1 1 0 0 0 0 0
27 : 0 0 1 1 0 1 1 0 0 0 1 0 0 0 1 0 0 1 0 0 1 1 0 0 0 1 1 0 0 1 1
28 : 1 0 0 1 1 1 0 1 0 1 1 0 1 0 1 1 0 1 1 1 1 1 1 0 1 0 0 0 1 0 1
29 : 1 1 0 1 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 0 1 0 0 1 1 1 1 0
30 : 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```
NUMBER OF 1's IN BOOLEAN MATRIX = 453

Fig. 3. Algorithm of constructing the Boolean matrix for a self-dual normal basis

Fig. 4. Boolean matrix (a) for $GF(2^9)$, (b) for $GF(2^{17})$, and (c) for $GF(2^{31})$
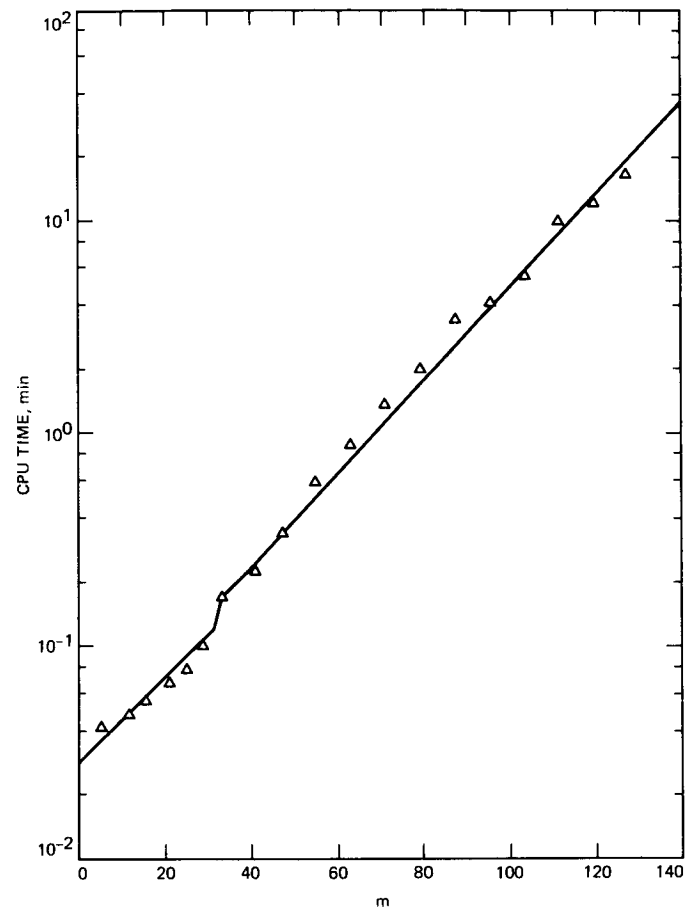
Fig. 5. CPU time required to construct Boolean matrix for $GF(2^m)$

Thermal Evaluation Method for Klystron RF Power

B. L. Conroy, H. Schleier, and T. Tesarek

Radio Frequency and Microwave Subsystems Section

This article examines the feasibility of adding instrumentation to the cooling system of a microwave transmitter for use as a calorimetric power measurement calibration. It considers the accuracy of the basic measurements as well as heat sources and losses not measured. Experimental results are presented in support of the theory.

I. Introduction

Figure 1 is a simplified block diagram of a transmitter, indicating the typical power measurement system. Power measured by the power meter depends on the coupling factor of the directional coupler, which typically can only be calibrated to an absolute precision of 0.5 dB or about 12%. A possible method of obtaining a more accurate absolute calibration is to measure the coolant flow to, and temperature rise across, the water load when the switch is in the water load position. By comparing the power measured in this way with the power meter reading, an in situ calibration factor can be obtained for the power meter for use when the switch is in the other position. This calibration factor includes the losses between the coupler and the switch as well as the coupler itself.

II. Dimensional Analysis

To calculate transmitter power from coolant measurements of flow rate and temperature rise, we need a K factor such that

$$\text{Power (kW)} = K \times \text{Flow (Gallons/Minute)} \times \Delta T \text{ (deg C)}$$

The dimensions of K must be (kW Minutes)/(Gallon deg C).

Handbooks and data sheets give the specific heat (Hs) of glycol mixtures in units of BTU/(lb deg F) and specific gravity (Gs) relative to water at 60 deg F as functions of percent glycol and temperature.

To reach the desired units we must go through the following chain of conversion factors:

$$K = Hs \frac{\text{BTU}}{\text{lb deg F}} \times Gs \frac{\text{g/cc}}{\text{g/cc @ 60}} \times 0.999 \text{ g/cc @ 60}$$

$$\times \frac{9 \text{ deg F}}{5 \text{ deg C}} \times 1.055 \frac{\text{kW sec}}{\text{BTU}} \times \frac{1 \text{ lb}}{453.59 \text{ g}} \times 3785 \frac{\text{cc}}{\text{gal}}$$

$$\times \frac{1 \text{ min}}{60 \text{ sec}}$$

$$K = 0.2638 \times Hs \times Gs \frac{\text{kW minutes}}{\text{gallon deg C}}$$

Thus to calculate the power it is necessary to know

(1) Flow rate

(2) Differential temperature

(3) Specific gravity

(4) Specific heat of the cooling fluid

The first two of these can be easily measured by a data acquisition system, but the second two are dependent on the composition of the fluid and temperature.

III. Measurement Accuracy

Volumetric Flow. A turbine flow meter for liquids can be obtained with a single point accuracy of 0.05% and linearity of 0.25% of the reading over a flow range of 10 to 1.

Temperature. Within the range of 0 to 100 deg C, a platinum RTD can be calibrated to an accuracy of ±0.03 deg C.

Fluid Density and Specific Heat. Typical cooling fluid is a mixture of deionized water with approximately 40% ethylene glycol by weight. The simplest measurement strategy is to draw a sample of the fluid, analyze it for percent glycol, and then calculate the density and specific heat. To facilitate this calculation, a multivariable, least squares curve fit produced the following expression for K, in terms of percent glycol (Pg) and degrees Celsius (Tc) which appears to be accurate to within 0.1% in the range of 20 to 40% glycol and 20 to 70 deg C:

$$K = 0.2651 - 7.048e\text{-}4\,Pg - 4.963e\text{-}6\,Pg^2$$

$$+ 5e\text{-}6\,Pg\,Tc + 6.417e\text{-}5\,Tc - 1.792e\text{-}6\,Tc^2$$

Differentiating this expression and evaluating around 40 deg C and 35 percent glycol yields:

$$(\Delta K)/K = -4.2e\text{-}3\,(\Delta Pg),$$

and

$$(\Delta K)/K = 3.1e\text{-}3\,(\Delta T)$$

Absolute accuracy of the measurement of percent glycol from a sample is around ±0.5.

Although the accuracy of an individual temperature measurement is 0.03 deg C, the fluid parameters vary with temperature through the range of ΔT. The proper solution is to integrate over temperature; however, a simpler strategy is to use the average temperature and a maximum error limit of about one fifth of the total temperature excursion, or about 2 deg C.

A. Potential Accuracy

By differentiating the log of the basic calorimetric expression it may be shown that:

$$(\Delta P)/P = (\Delta K)/K + (\Delta\ \text{flow})/\text{flow} + \Delta(\Delta T)/(\Delta T)$$

Using

$$(\Delta Pg) = 1\%$$
$$(\Delta Tc) = 2\ \text{deg C}$$
$$(\Delta K)/K = 4.2e\text{-}3 \times (\Delta Pg)$$
$$+ 3.1e\text{-}3 \times (\Delta Tc) = 0.0104$$
$$(\Delta\ \text{flow})/\text{flow} = 0.0025$$
$$(\Delta(\Delta T))/(\Delta T) = 2 \times 0.03/10 = 0.006$$

yields a potential inaccuracy of

$$(\Delta P)/P = 0.0104 + 0.0025 + 0.006$$

$$= 0.0199\ \text{or}\ 2\%$$

B. Other Errors

Two other limitations of the accuracy of a calorimetric power measurement are:

(1) Heat is generated in the fluid as it flows, due to viscous losses.

(2) Heat is lost to the environment due to conduction, convection and radiation, without being transferred to the cooling fluid.

1. Viscous losses. Assume a water-like liquid flowing at 11 gpm in a 1 and 1/2-in. SCH40 smooth pipe.

D (diameter) = 1.61 in./12 in./ft = 0.134 ft

$\overline{V}$ (velocity) = 11 gpm/6.34 gpm/fps = 1.74 ft/s or 6246 ft/h

ρ (density) = 62.4 lb_m/ft^3

α (viscosity) = 1 cp × 2.42 lb_m/ft/h/cp

= 2.42 lb/ft/h

R_e (Reynolds No.) = $D\overline{V}\rho/\alpha$

$$= \frac{0.134\ \text{ft} \times 6246\ \text{ft/h} \times 62.4\ lb_m/\text{ft}^3}{2.42\ lb_m/\text{ft/h}}$$

= 21581 or f = 0.006

$$H_f = \frac{4 f \bar{V}^2 L}{2 gcD}$$

$$= \frac{4 \times 0.006 \times (1.74 \text{ ft/s})^2 \times 1 \text{ ft}}{2 \times 32.2 \dfrac{\text{ft lb}_m}{\text{lb}_f - s^2} \times 0.134 \text{ ft}}$$

$$= 8.42\text{E-}3 \text{ ft/lb}_f/\text{lb}_m$$

$$P = 8.42\text{E-}3 \frac{\text{ft/lb}_f}{\text{lb}_m} \times (500 \times 11) \frac{\text{lb}_m}{\text{h}} \times 3.777\text{E-}7 \frac{\text{kWh}}{\text{ft/lb}_f}$$

$$= 1.75\text{E-}5 \text{ kW/ft of pipe}$$

If the length between measurement points across the water load are within 1 foot and the total power is on the order of 10 kW then:

$$(\Delta P)/P = 1.75\text{e-}5/10 = 1.75\text{e-}6 \text{ or } 0.00002\%$$

2. Thermal losses. Assume a condition of 50°C heat loss to surrounding at 0°C for both radiation and conduction-convection with a loss of 2 B/ft²/h°R for conduction-convection, and for radiation a black-body in large surroundings.

$$T_H = 50°\text{C} = 323 \text{ K} = 581°\text{R}$$

$$T_C = 0°\text{C} = 273 \text{ K} = 491°\text{R}$$

and

$$g/A = 1728\text{E-}12 \frac{B}{\text{ft}^2/\text{h}/°\text{R}^4} [(581°\text{R})^4 - (491°\text{R})^4]$$

$$+ \frac{B}{\text{ft}^2/\text{h}°\text{R}} [581°\text{R} - 491°\text{R}] = 276 \frac{B}{\text{ft}^2/\text{h}}$$

or

$$276 \frac{B}{\text{ft}^2/\text{h}} \times 0.00293 \frac{\text{kWh}}{B} = 0.08 \text{ kW/ft}^2$$

Using a typical area of 0.1 ft² for the water load, the error due to this source is:

$$(\Delta P)/P = 0.008/10 = 0.0008 \text{ or about } 0.1\%$$

Conducted losses can further be reduced by control of the cooling flow to the waveguide next to the water load so that the waveguide is at the same temperature as the load.

IV. Experimental Results

Several test runs were made on a 20 kW X-band (7.190 GHz) klystron, comparing the DC power calculated by the product of the beam voltage and current (with no drive) to the power calculated from the flow and ΔT across the tube collector. Figure 2 is a graph of the measured collector power vs. time. Figure 3 shows the temperature of the coolant entering and leaving the collector, and Fig. 4 is a detail of the percent difference between the two measurement methods. Although the absolute calibration of the DC power measurement is no better than 2%, two important features may be noted:

(1) A change in beam power introduces a transient error in the thermal measurement until the temperature stabilizes.

(2) Even when the beam power is constant, the error still shows some residual temperature effects. In the period from 15 to 23 minutes while the DC power is constant at about 52 kW, the inlet temperature rises about 5 deg C and there is a small drift in the percent difference. At low DC power and with large temperature differences, the effect is larger. While the periods just before 5 minutes and just after 25 minutes are both at about 7 kW DC, the inlet temperature has increased by about 15 deg C and the change in percent difference is over 2%.

V. Conclusions

(1) Thermal measurements offer a theoretical potential for a substantial increase in the accuracy of RF power calibration.

(2) Additional benefits can be achieved by better modeling of the fluid, particularly by integrating the specific heat over the temperature span.
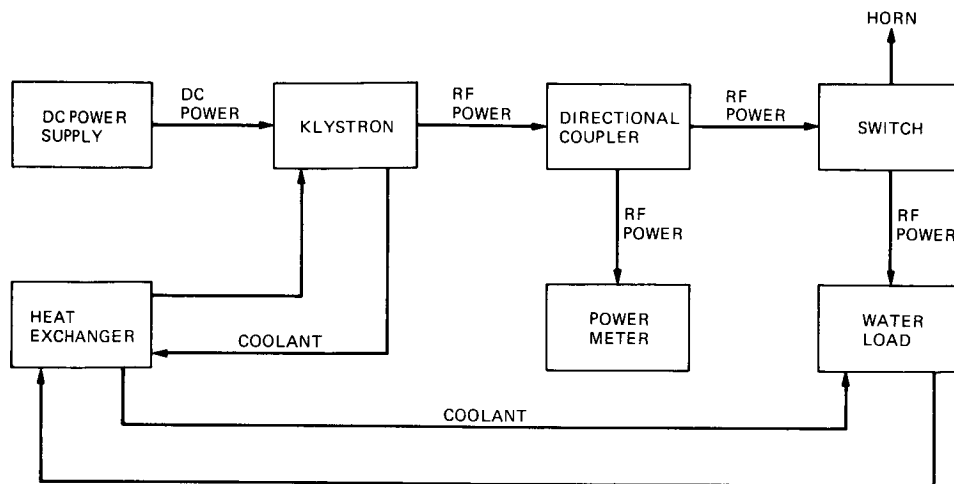
Fig. 1. Simplified transmitter block diagram
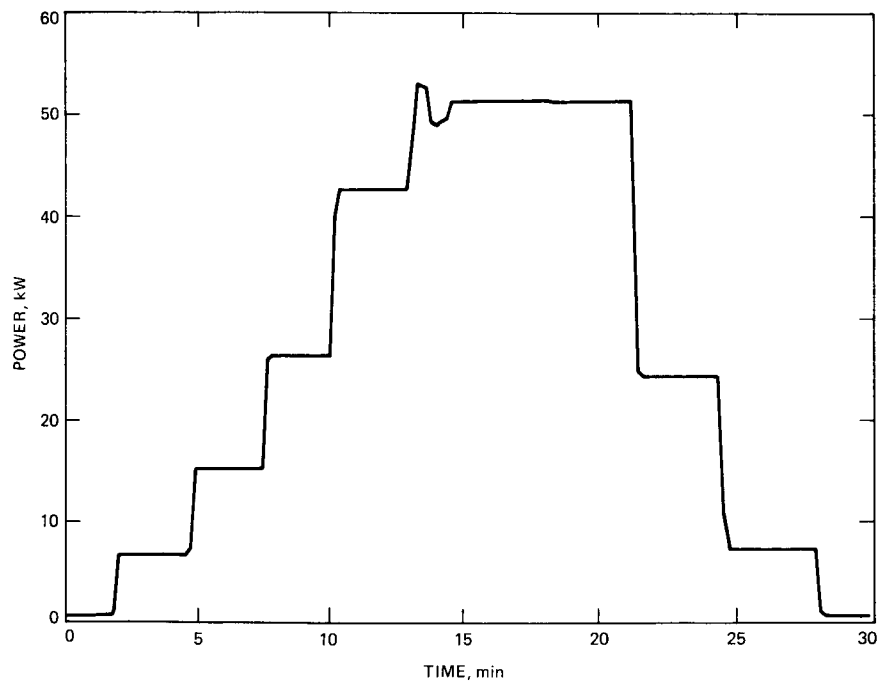
Fig. 2. DC beam power vs time
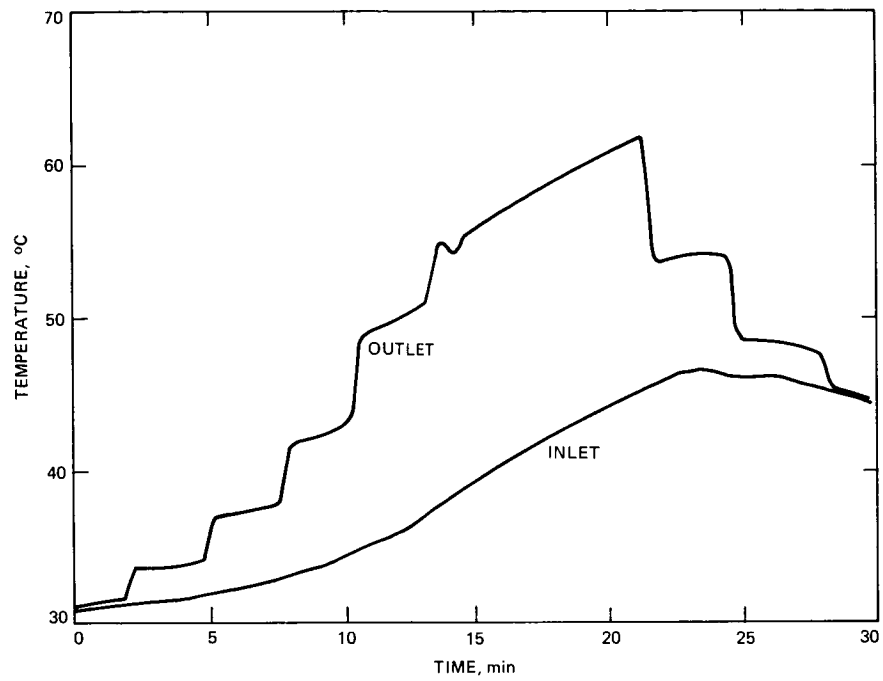
Fig. 3. Inlet and outlet coolant temperature
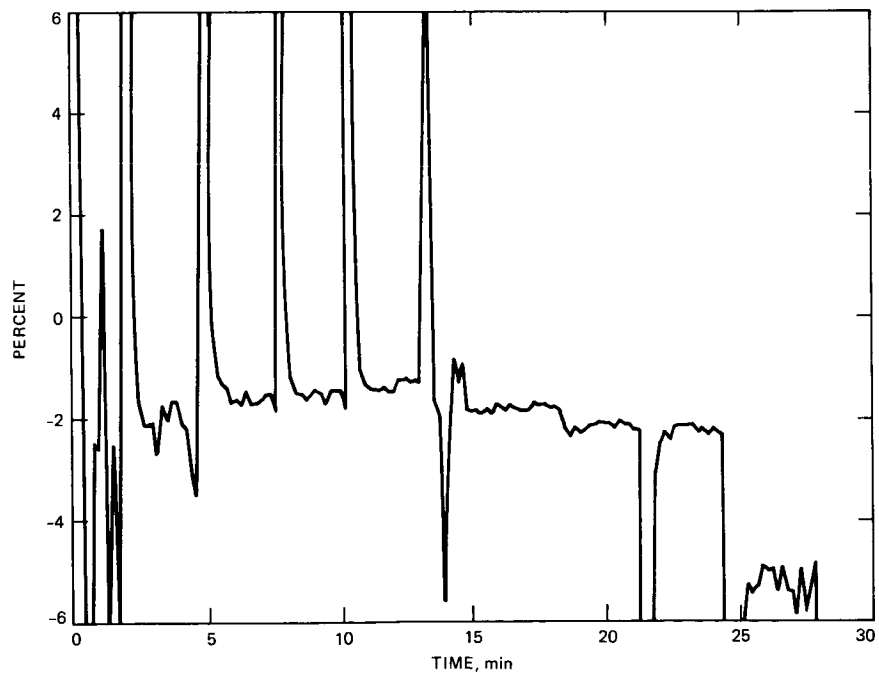
Fig. 4. Percent difference between DC measurement and thermal measurement

Ka-Band (32-GHz) Downlink Capability for Deep Space Communications

J. G. Smith

TDA Technology Development Office

The first quarter century of U.S. solar system exploration using unmanned spacecraft has involved progressively higher operating frequencies for deep space telemetry: L-band (960 MHz) in 1962 to S-band (2.3 GHz) in 1964 to X-band (8.4 GHz) in 1977. The next logical frequency band to develop for deep space is Ka-band (32 GHz) for which a primary deep space allocation of 500 MHz between 31.8 GHz and 32.3 GHz was established in 1979. The telecommunications capability was improved by a factor of 77 (18.9 dB) through the frequency changes from L-band to X-band. Another improvement factor of 14.5 (11.6 dB) can be achieved by going to Ka-band.

Plans to develop and demonstrate Ka-band capability include the continued measurement of weather effects at DSN sites, development of a prototype DSN ground antenna and supporting subsystems, augmentation of planned spacecraft with Ka-band beacons, and development of spacecraft prototype modules for future Ka-band transmitters. Plans for augmenting the DSN with Ka-band capability by 1995 have also been developed.

A companion set of articles in this issue describes the Ka-band performance and technology in greater detail.

I. Introduction

The exploration of the solar system with spacecraft has expanded beyond the narrow confines of Earth from the early Pioneers and Mariners to the presently on-going odysseys of the Voyagers and later Pioneers well beyond two billion miles from Earth. With the exception of the Mariner '71, Mars Orbiter, Viking, and Pioneer Venus, these missions were fly-bys intended primarily for early reconnaissance. Missions planned for the near future include largely orbiters: Galileo in orbit around Jupiter; Mars Observer around Mars; and Magellan around Venus. Subsequent missions will be of increasing complexity: rendezvous, landing, roving, and return.

Spacecraft for these missions face stringent constraints: continued high performance on more complex missions with only a modest increase of mission investment. Fortunately, new spacecraft and ground capabilities continue to help ease these constraints. Within the domain of deep space communi-

cations, there has been a millionfold increase, for example, in the telemetry capability.

Figure 1 shows a profile of the DSN's capability to support the reception of increasingly greater data rates from increasingly greater distances. The abscissa shows the calendar years during which particular increases in capability first appeared. The ordinate shows the data rate that could have been supported in that year by the Deep Space Network (DSN) if the data had been sent from spacecraft at Mars or Neptune. The actual data rates that could have been supported by the DSN from a spacecraft at any particular distance scales inversely as the square of the relative distances.

Major jumps in the staircase function of Fig. 1 occurred as the frequency of operation increased from L-band (0.96 GHz) to S-band (2.3 GHz) to X-band (8.4 GHz). This occurs primarily because the capacity of a link established between two well-aimed high-gain antennas scales, to a first order, as the square of the operating frequency. Thus, moving from 0.96 GHz to 2.3 GHz offered a possible increase of 574% (or 7.6 dB), and from 2.3 GHz to 4 GHz another factor of 13.5 or 11.3 dB. Actual gains are reduced by propagating media effects and the limiting tolerances of the various components in the link, but may be increased by the reconfigurations possible at higher frequencies.

Command frequencies for uplink communication are also increased as telemetry frequencies increase, although usually later. This tendency to follow is because lower frequency uplinks multiplied up to the telemetry frequency in the spacecraft transponder introduce unwanted jitter in the downlink. The deep space telemetry frequency was shifted to X-band in 1977, but the command frequency is only now shifting to X-band. Figure 2 shows the trend to higher frequencies, including a projected increase to Ka-band (32 GHz for telemetry and 34 GHz for command).

The frequencies used for deep space communication tend to be those for which protection against interference exists. Figure 3 shows the present allocations available on a primary (protected) basis (Ref. 1). The next major frequency increase available for deep space is the move from X-band to Ka-band (32 GHz). The potential increase in link capability is 11.6 dB, although the actual increase is expected to be slightly less than that because of atmospheric effects and the performance of the antennas of the DSN that were designed for good performance at the lower X-band frequency. The use of quasi-optical techniques, and reduced spacecraft antenna diffraction losses, can offset these losses.

Figures 4 and 5 show in two ways why Ka-band was selected—it is basically the last good location in the microwave

"window" (Fig. 4) and offers, at least in clear, dry weather, an increase of 10 dB (Fig. 5). As the series of companion articles following shows, the strong potential exists for 8 dB to 10 dB improvement even with the limitations imposed by weather (Refs. 2, 3, 4, 5). The benefits of this potential to specific planned missions are also under study (Ref. 6).

The impact of only a 6 dB increase is reflected in Fig. 6. This shows, first, that a very large array of X-band antennas has the same reception capability (G/T) as one 70-meter antenna at Ka-band and, second, that a 70-meter X-band antenna has the same reception capability as one 34-meter antenna at Ka-band.

II. Historical Perspective

The consideration of Ka-band for deep space communication began more than a decade ago, in a variety of advanced mission planning studies. One of the first formal treatments occurred in Hunter's 1976 study "Orbiting Deep Space Relay Station" (Ref. 7). One of the outcomes of the 1979 World Administrative Radio Conference was the primary allocation for deep space communication only, in the three countries of interest, U.S.A., Australia, and Spain, with 31.8 to 32.3 GHz for space-to-ground and 34.2 to 34.7 GHz for ground-to-space (Ref. 1).

The argument for a ground-based Ka-band receiving station was strengthened by the realization in 1981 through measurement of the weather effects at 32 GHz (Ref. 8) that Ka-band reception was not as badly degraded by weather effects as previously anticipated.[1] Further, studies showed that DSN antennas could be upgraded to perform well at Ka-band.

The selection of Ka-band as the next choice to develop for deep space telemetry was not easy or simple. The belief still existed on the one hand that the combination of weather and DSN antenna limitations precluded meaningful increases in telemetry capability at higher frequencies, and that deep space communications should therefore remain indefinitely at X-band. On the other hand, it was argued that the shift to Ka-band would be too modest, and that a shift to an optical band for deep space communication would be more appropriate. The difficulty of resolving this issue with all involved elements at JPL and at NASA Headquarters delayed for a few years funding for the technology development necessary to enable Ka-band communications. References 9 and 10 show some of

---

[1] Letter from W. H. Bayley, Jet Propulsion Laboratory, Pasadena, California, to H. G. Kimball, NASA Headquarters, Code TN, "Applicability of the 32 GHz Frequency Region to Deep Space Communications," July 9, 1979 (internal document).

the trade-offs produced that finally enabled the decision to be resolved.

Thus, Ka-band has been accepted as the next logical frequency to develop capability for deep space communication, with optical as a reasonable follow-on. This decision became involved with a proposal for a new R&D antenna for the DSN (Ref. 11), a proposal for a Ka-band spacecraft beacon experiment (Ref. 12), and proposals for development of Ka-band components for future missions. Specifically, JPL concluded the following:

(1) Ka-band for deep space missions appears to be cost effective and to have an extended useful operational lifetime regardless of exactly when it is introduced.

(2) JPL should continue with serious Ka-band studies, making decisions as appropriate, but specifically should work toward including a Ka-band beacon on Mars Observer.

(3) A new R&D antenna should be included in the FY88 Construction of Facilities budget. (There is major potential synergism with Ka-band activities; the rationale and need for an R&D antenna at Goldstone are independent of the Ka-band decision.)

(4) A full-up operational optical communication system for the deep space network is not realistic before 2010.

(5) JPL should continue an appropriate level of activity, including monitoring the development of optical communications technology, doing deep space related specialized optical communication technology development, and considering the system implications of the use of optical communication for deep space missions.[2]

Subsequently, a Notice of Intent was signed, confirming agreement of both JPL and NASA Headquarters (Office of Aeronautics and Space Technology, Office of Space Tracking and Data Systems, and Office of Space Science and Applications) to proceed with the key ingredients to develop the Ka-band technology for subsequent deep space communications.

III. Current Efforts

The beginning of the conversion of deep space communication to Ka-band will not occur until the Ka-band technology has been developed and demonstrated to provide the benefits anticipated and modeled. Several steps are involved:

(1) Continued measurement of weather effects at Ka-band at all three DSN complexes (U.S., Australia, and

Spain), with water vapor radiometers to refine the weather model essential to projections of deep space link performance.

(2) Development of a DSN R&D ground antenna to verify the benefits of various approaches to be used to upgrade the existing 34-meter and 64/70-meter antennas to good efficiency at Ka-band. This is planned to be in operation by 1990.

(3) Development of a prototype low-noise Ka-band traveling wave maser of suitable gain-bandwidth product to ensure low system noise temperature performance at this higher frequency. Also, development of pointing capability consistent with Ka-band.

(4) Augmentation of planned deep space spacecraft with Ka-band beacons to enable simultaneous X-band and Ka-band reception in order to obtain accurate relative performance measurements. Use of X-band fourth harmonics is expected to permit the addition of experimental beacons at modest cost.

(5) Development of spacecraft modules to provide 2 to 50 Watts of Ka-band transmitted power at acceptable DC-to-RF efficiency. Also, provision for vernier pointing of the spacecraft beam over a few beamwidths. Figure 7 shows a possible progression of modules to be developed: the NASA X-band transponder (NXT) now under development by Cubic Corporation, X-band solid state power amplifiers (XSSPA) under development at JPL, a Ka-band exciter (KEX) to upconvert the X-band frequency of the NXT, a dual frequency X-band/Ka-band feed arrangement (XKFA) shown in the figure as a dichroic plate with separate feeds, and a phased array of Ka-band elements.

IV. A Plan for Operational Use by 1995

Once Ka-band has been demonstrated in the field through experiments with spacecraft and the DSN R&D station, conversion of new spacecraft and augmentation of the DSN to accommodate the new frequency will begin. Flight projects will rely on the developed prototypes and flight components to produce flight qualified elements for subsequent flights.

The augmentation of Ka-band to the DSN requires three distinct steps: modification of the 70-meter antennas, and the 34-meter high efficiency antennas (HEFs), and replacement of the 34-meter standard (STD) antennas. There is no plan to convert the 26-meter subnet. These steps will all be simplified if the DSN antennas are converted to beam waveguide operation as presently envisioned (Ref. 13, 14, 15).

---

[2] R. J. Parks, unpublished communication, April 16, 1986.

The 70-meter antennas after upgrade from 64 meters are expected to be usable at Ka-band with about 35% efficiency. This is based on data relative to current performance at 22 GHz and specifications established for the upgrade. Increased efficiency is expected to be possible through array feeds.

The 34-meter high efficiency (HEF) antennas are already expected to be 50% efficient at Ka-band according to calculations based on surface tolerances. Conversion to a beam waveguide configuration would again simplify addition of Ka-band telemetry capability because of added space for electronics.

A possible schedule for augmentation of the deep space communication channel to include Ka-band is shown in Fig. 8. This shows conversion of DSS 14 by the end of 1994 to be used for Cassini (Saturn Orbiter/Titan Probe) at encounter in 2001. The upgrade of the rest of the DSN would occur during the long Cassini cruise. Actual conversion would depend on other deep space missions of that era not yet well defined (e.g., Mars Sample Rover).

V. Conclusion

The augmentation of the deep space telemetry channel with Ka-band can be thought of as having started in 1979 when the deep space allocations were obtained. It will probably be completed in the mid-90s as all three DSN subnets are converted to provide a Ka-band capability and the deep space spacecraft all

utilize Ka-band as their primary telemetry frequency in order to keep down mission costs or enhance mission performance.

The following articles on Ka-band provide details on the steps taken so far to provide deep space missions with the advantages of a Ka-band telemetry capability. They are not always consistent one with another, generally because they represent different stages of an evolving understanding of Ka-band and how it will eventually be implemented for deep space communication.

These articles are grouped into three categories: general, performance, and future activities. Within the general category, de Groot defines the Ka-band allocation (Ref. 1); Hansen and Kliore identify the benefits to specific deep space missions (Ref. 6); and Layland and Smith map out a possible scenario for evolving not only to Ka-band, but beyond to optical frequencies (Ref. 10). Next are articles on Ka-band performance of 70-meter antennas by Bhanji, et al. (Ref. 3); and weather effects at Ka-band by Slobin (Ref. 5). Future activities include a Ka-band beacon experiment by Riley (Ref. 12); measurements of Galileo harmonics for another Ka-band beacon experiment by Stanton and Manshadi (Ref. 16); a planned new DSN research and development antenna to receive the Ka-band beacon signals by Smith (Ref. 11); a plan by Riley to develop Ka-band spacecraft systems (Ref. 4); the benefits of beam waveguides in the DSN and their benefits to Ka-band operation of the DSN, by Clauss and Smith (Ref. 15); and some beam waveguide structural design considerations by Katow, et al. (Ref. 16).

References

1. de Groot, N. F., "Ka-band (32 GHz) Allocations for Deep Space," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

2. Koerner, M. A., "Relative Performance of X-band and 32 GHz Telemetry Links on the Basis of Total Data Return Per Pass," *TDA Progress Report 42-87*, Jet Propulsion Laboratory, Pasadena, California, pp. 65-80, November 1987.

3. Imbriale, W. A., Bhanji, A. M., Blank, S., Lobb, V. B., Levy, R., and Rocci, S., "Ka-band (32 GHz) Performance of 70-meter Antennas in the Deep Space Network," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

4. Riley, A. L., "Ka-band (32 GHz) Spacecraft Development Plan," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

5. Slobin, S. D., "Models of Weather Effects on Noise Temperature and Attenuation for Ka-band and X-band Telemetry Performance Analysis," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, Calif., February 1987.

6. Hansen, D. M., and Kliore, A. J., "Ka-band (32 GHz) Benefits to Planned Missions," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

7. Hunter, J. A., et al., "Orbiting Deep Space Relay Station Study, Final Report," *JPL Publication 79-30*, Jet Propulsion Laboratory, Pasadena, California, June 15, 1979.

8. Slobin, S. D., Franco, M. M., and Stelzried, C. T., "20.7- and 31.4-GHz Atmospheric Noise Temperature Measurements," *TDA Progress Report 42-64*, Jet Propulsion Laboratory, Pasadena, California, pp. 132-139, May-June 1981.

9. Dickinson, R. M., "Comparison of 8.415-, 32.0- and 565645-GHz Deep Space Telemetry Links," *JPL Publication 85-71*, Jet Propulsion Laboratory, Pasadena, California, October 15, 1985.

10. Layland, J. W., and Smith, J. G., "A Growth Path for Deep Space Network Communications," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

11. Smith, J. G., "Proposed Upgrade of the Deep Space Network Research and Development Station," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

12 Riley, A. L., Hansen, D. M., Hartop, R. W., and Mileant, A., "A Ka-band Beacon Link Experiment (KABLE) with Mars Observer," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

13. Clauss, R. C., and Smith, J. G., "Beam Waveguides in the Deep Space Network," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

14. Veruttipong, T., Withington, J., Galindo-Israel, V., Imbriale, W., and Bathker, D., "Design Considerations for Beam Waveguide Antennas in the Deep Space Network," *TDA Progress Report 42-87*, Jet Propulsion Laboratory, Pasadena, California, pp. 1-9, November 1987.

15. Katow, S., Cucchissi, J. J., Chuang, K. L., Levy, R., Lansing, F. L., Stoller, F. W., and Menninger, F. J., "Structural Design Options for the New 34-meter Waveguide Antenna," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

16. Stanton, P. H., and Manshadi, F., "Harmonic Measurements of the Galileo Spacecraft X-band Transmitter System," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

Fig. 1. Profile of deep space telemetry capability
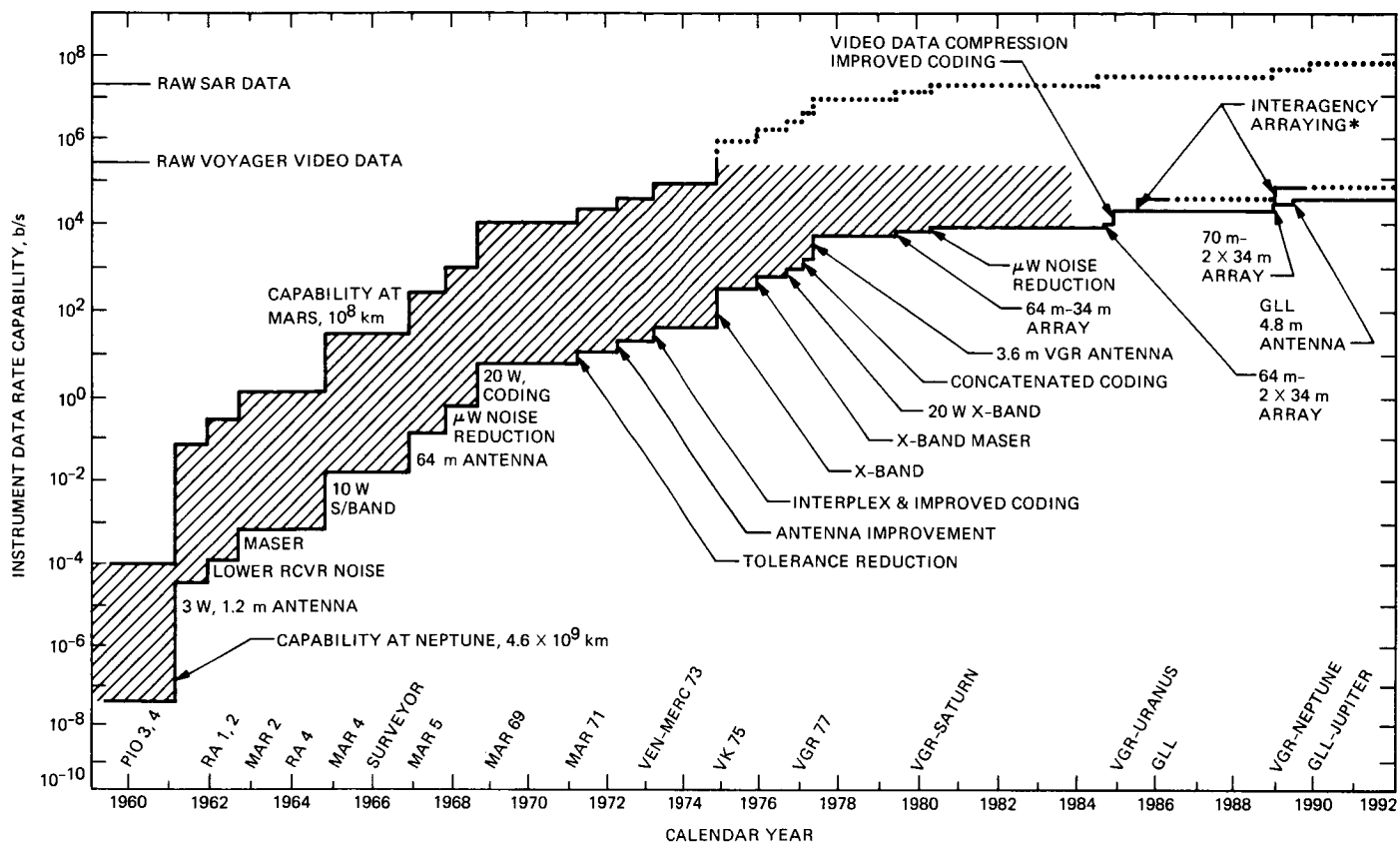
* INTER-AGENCY ARRAYING IS USED ONLY FOR SPECIAL EVENTS
(e.g., VGR AT URANUS - NEPTUNE)
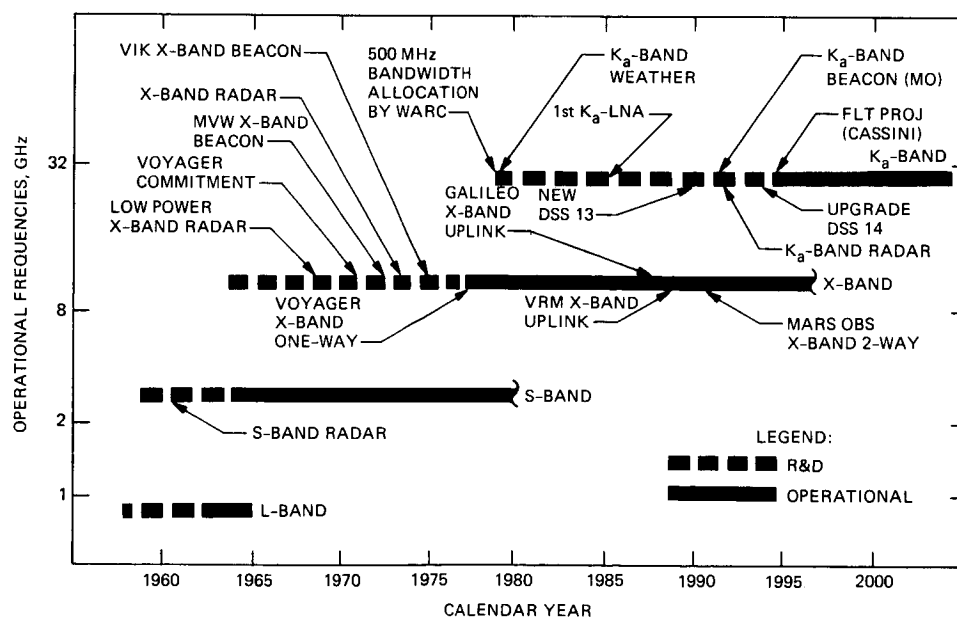
Fig. 2. Progression of deep space frequency bands

SPACECRAFT TO EARTH (DOWNLINK)
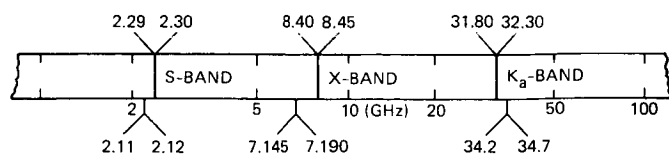
EARTH TO SPACECRAFT (UPLINK)

Fig. 3. The 1979 allocation at Ka-band to augment existing bands
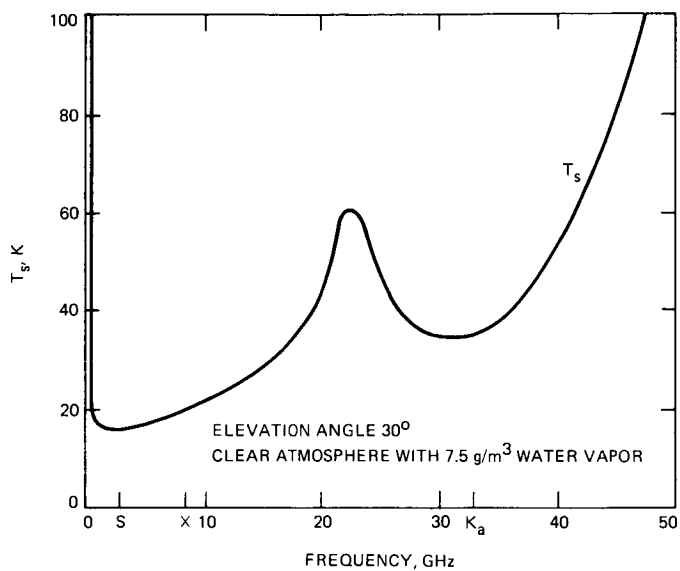
Fig. 4. Projected DSN 70-meter antenna system temperature
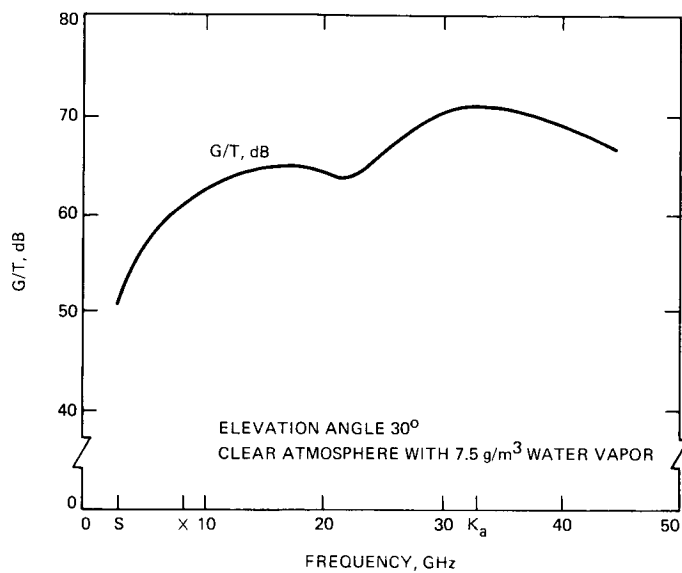
Fig. 5. Projected DSN 70-meter antenna system G/T
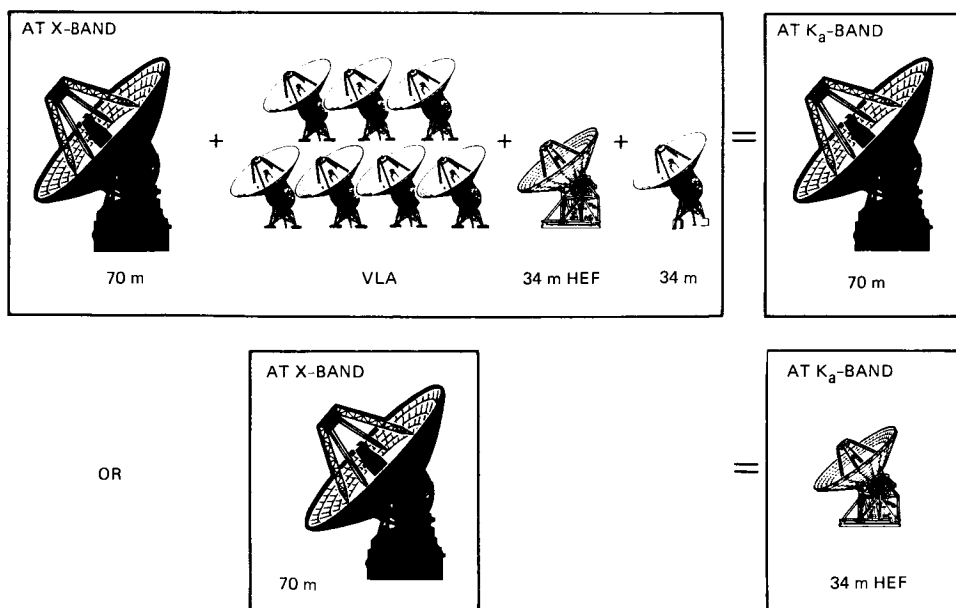
Fig. 6. The X- and Ka-band G/T equivalences
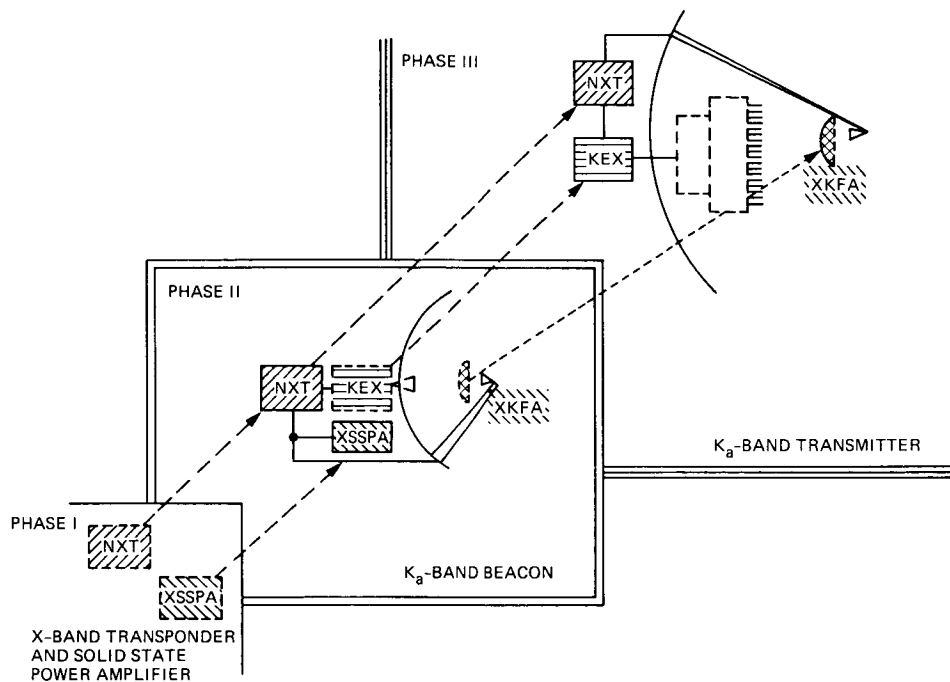
Fig. 7. Modular additions of spacecraft hardware
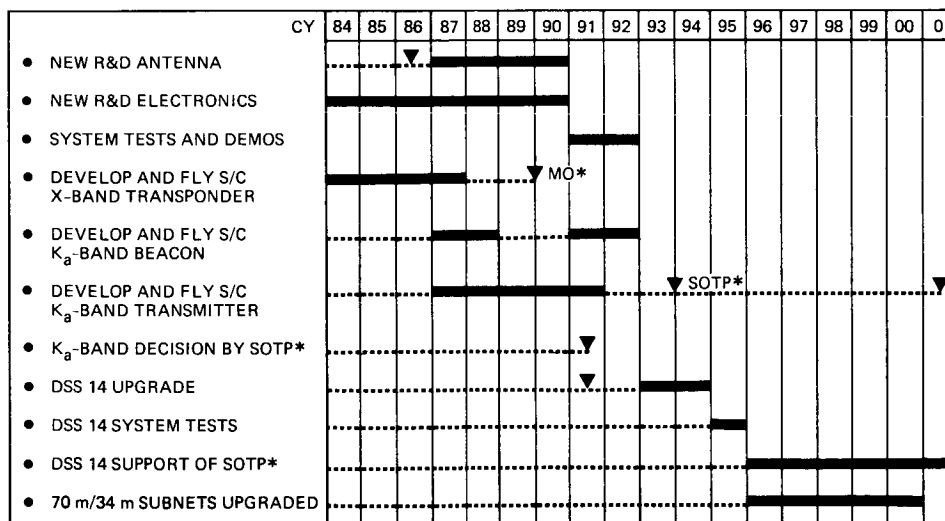
*MO - MARS OBSERVER
SOTP - SATURN ORBITER/TITAN PROBE

Fig. 8. Plan for Ka-band upgrade

TDA Progress Report 42-88

Ka-Band (32 GHz) Allocations for Deep Space

N. F. de Groot

TDA Technology Development Office

At the 1979 World Administrative Conference, two new bands were allocated for deep space telecommunications: 31.8–32.3 GHz, space-to-Earth, and 34.2–34.7 GHz, Earth-to-space. These new bands provide opportunity for further development of the Deep Space Network and its support of deep space research.

This article discusses the history of the process by which JPL/NASA developed the rationale, technical background, and statement of requirement for the new bands. Based on this work, United States proposals to the conference included the new bands, and subsequent U.S. and NASA participation in the conference led to successful allocations for deep space telecommunications in the 30 GHz region of the spectrum. A detailed description of the allocations is included in the article.

I. Introduction

Successful deep space telecommunication requires some international and national basis for use of specific frequency bands with provision for sharing with other users, and provision for providing protection of individual channels from interference from other users.

This article presents a brief history of steps that led to the international allocation of new bands for deep space telecommunications in the 30 GHz region of the radio frequency spectrum. Following the historical narrative, a rather detailed description of the allocations is given.

II. The 32/34 GHz Band Allocations for Deep Space Research: How They Were Obtained

In 1977, the United States was preparing for the 1979 World Administrative Radio Conference (WARC). At this conference, the entire radio frequency spectrum and its usage as governed by the Radio Regulations was to be reconsidered. The result of the WARC would be a new treaty, adopted by 150 nations of the International Telecommunication Union (ITU). Such a broad ranging conference is rare, and decisions taken in 1979 would likely affect the use of much of the radio spectrum until the turn of the century.

NASA was a major participant in the U.S. preparations for WARC 79, looking at the need for existing and new radio frequency bands to support the agency's future programs. At that time, JPL studies suggested that an orbiting deep space relay satellite (ODSRS) could be a next step in developing a greater capability for the DSN. The satellite would be in a geostationary orbit, and provide communication links to spacecraft in deep space and to the Earth (Ref. 1).

To protect the precious deep space bands, it was proposed to use other bands for the links between the Earth and the

satellite. For the links between the satellite and deep space, three candidate frequency ranges were identified. A pair of bands near 31 GHz would provide for links to and from deep space, but with the advantage of also permitting links directly with DSN stations on Earth. This possibility arises from the microwave propagation window through the atmosphere near that frequency. The ability to perform a deep space mission from either the ODSRS or the Earth would provide a safer environment for mission communications during development of the satellite capability. A band pair near 95 GHz would also allow a degree of direct communication with earth stations, although development for links at such a high frequency would be far in the future.

Ultimately, the satellite to deep space links would have the best protection from terrestrial radio interference if they were in a part of the spectrum where the atmosphere is relatively opaque. A pair of bands near 60 GHz would provide high atmospheric attenuation and the desired protection.

In mid 1977, it was proposed by JPL[1] that the U.S. should seek deep space allocations for uplink-downlink band pairs near 31, 60, and 95 GHz in support of the future Orbiting Deep Space Relay Satellite (ODSRS). Each band was to have primary status, be 500 MHz wide, and have a specified direction of transmission.

In reply, NASA[2] asked that a description of ODSRS frequency requirements be prepared for use in U.S. WARC deliberations, and that a technical paper be prepared for submission to the Comite' Consultatif International des Radiocommunications (CCIR). That organization was scheduled to hold a Special Preparatory Meeting (SPM) in the fall of 1978 to lay the technical groundwork for the 1979 administrative conference. CCIR is the technical arm of the ITU with regard to radio communications.

In early January 1978, a draft CCIR technical report[3] was sent to the chairman of U.S. CCIR Study Group 2. This group deals with space research and radio astronomy. The paper provided the rationale for the choice of ODSRS frequency bands, and was submitted as an input to the CCIR process of preparing for its SPM in support of the conference.

The CCIR draft report was also forwarded to NASA[4] for its review, but with an additional proposal that six specific bands (three uplink-downlink pairs) be sought by the United States for ODSRS purposes.

A week later, the list of NASA frequency requirements proposed for inclusion in the U.S. submission to the 1979 conference was distributed.[5] The list did not include the bands and status proposed by JPL for ODSRS links. Provision for ODSRS was evidently held by NASA to be insufficient justification for protected deep space links above 20 GHz.

The effort to have technical input to the CCIR failed. The CCIR paper concerning frequencies for ODSRS did not leave the United States. Not having input from the U.S., the CCIR SPM held in October–November 1978 did not deal with the issue of new deep space bands above 20 GHz. Consideration of the matter by the WARC in 1979 would therefore have to rest on U.S. proposals to the conference itself, and not CCIR technical advice coming from the SPM.

During 1978, the U.S. proposals for the WARC became more fully developed, and included the NASA list of requirements, essentially as presented in January.

The situation in February 1979 appeared to be that the then current U.S. proposal for radio frequency allocations foreclosed the option of protected deep space telecommunications links in the 20–120 GHz region, except for a band at 65–66 GHz. Allocations proposed at 31–31.3 GHz and 31.8–32.3 GHz would permit deep space links, but without any protection from interference.

Although it was late, JPL undertook the task of getting the U.S. position changed. It was clear that a stronger rationale than ODSRS alone was needed. A number of steps were taken.

(a) In early February, 1979, five reasons for maintaining the option to operate a ground-based DSN near 30 GHz were stated:

(1) 30 GHz is a candidate technology in the evolution toward increased link performance as needed to enhance or enable future missions;

---

[1] de Groot, N. F. (JPL letter 339-77-128; internal document), Jet Propulsion Laboratory, Pasadena, California, 13 June 1977.

[2] Eaton, E. L. (NASA letter TN/#1483; internal document), Jet Propulsion Laboratory, Pasadena, California, 25 August 1977.

[3] de Groot, N. F. (JPL letter 3398-78-253; internal document), Jet Propulsion Laboratory, Pasadena, California, 3 January 1978.

[4] Goodwin, P. S. and Bayley, W. H. (JPL letter 420-PSG:mti; internal document), Jet Propulsion Laboratory, Pasadena, California, 11 January 1978.

[5] Associate Administrator for Space Tracking and Data Systems, *NASA Frequency Requirements for the General World Administrative Radio Conference 1979 (GWARC 1979)* (NASA letter TN/1607; internal document), Jet Propulsion Laboratory, Pasadena, California, 16 January 1978.

(2) 30 GHz is the next logical step in the evolution of technology development (from 2 to 8 to 30 GHz to submillimeter to optical, rather than an intermediate step at 13 GHz);

(3) 30 GHz is the next logical step in the evolution of deep space allocations from the congested lower frequency bands near 2 and 8 GHz to broader, less congested higher frequency bands;

(4) 30 GHz offers potential sharing of technology developments with other near-Earth NASA sponsors (especially OAST and OSSA); and

(5) 30 GHz is the next logical step in the evolution from a ground-based Deep Space Network (DSN), with a "clear weather" window to back up an ODSRS.

(b) JPL emphasized to NASA[6] its future needs and the likely evolution of technology and the DSN, and asked that NASA reconsider its needs in connection with preparations for the WARC.

(c) The feasibility of incorporating the necessary frequency multipliers and dividers in a spacecraft transponder to operate near 30 GHz was confirmed. The rationale for the new bands was further developed. Advantages to flight projects and radio science were solidified. JPL encouraged NASA support for new deep space bands[7] and informed NASA[8] that a more fully developed analysis of the need for 30 GHz was being documented, along with a recommendation for specific band pairs to be proposed at the WARC.

In mid-May, a lengthy study on the applicability of the 32 GHz frequency region to deep space telecommunications was completed at JPL. The study considered attenuation, noise temperature, wavefront distortion, and related aspects of propagation through space and the atmosphere. Also considered were the effects of antenna surface precision, pointing accuracy, and spacecraft transmitter efficiency. These factors were all treated in a numerical comparison of estimated link performance at 8 GHz and 32 GHz. The potential advantage of 32 GHz deep space links was evident, amounting to an estimated maximum performance increase of 6.8 dB, as compared to 8.4 GHz, assuming 90 percent weather confidence and a 30 degree elevation angle. Subsequent studies have shown even greater increases (Ref. 2).

In early July, an edited version of the JPL position was sent to NASA.[9] A discussion of benefits to radio science was included, along with a list of four candidate band pairs in the 30.9 to 35.1 GHz region. The letter asked that all necessary steps be taken by NASA to secure the needed band allocations. In light of expected resistance, a request for bands near 60 and 100 GHz was not included.

Just eight days later NASA wrote to the U.S. WARC delegation spokesman for the Allocations Committee dealing with the 1215 MHz to 40 GHz frequency range, and asked that the JPL/NASA requirement for new bands near 32 GHz be included in the U.S. proposal to the WARC.[10] The requested change to the U.S. conference proposal was considered and adopted. The formal U.S. position concerning bands for deep space research had been successfully changed, albeit at the last minute.

In final preparation for the conference, JPL advised NASA[11] as to various negotiating trade-offs that might be needed in guiding action by the U.S. delegation at the conference. The WARC began its ten week deliberations in late September 1979. As it turned out, the principal obstacle to the requested deep space allocations was potential non-NASA use of that frequency within the U.S. This problem was resolved by a memorandum of understanding that geographically restricts 32/34 GHz deep space operations within the U.S. to Goldstone, California.

In its new Table of Allocations, the 1979 General WARC included two new bands for deep space research near 30 GHz. The lengthy and sometimes intense activity by JPL and NASA had resulted in new opportunities for deep space exploration, and provided the basis for further evolution of the DSN capability to serve flight projects.

III. The 32/34 GHz Allocations for Deep Space Research: What They Are

Radio frequency usage is based on band allocations by the ITU. A Table of Allocations is a major part of the Radio Regulations (Ref. 3). The Regulations are the expression of an

[6]Bayley, W. H. (JPL letter 410-WHB:baw; internal document), Jet Propulsion Laboratory, Pasadena, California, 6 March 1979.

[7]James, J. N. (JPL letter 910-JNJ:pd; internal document), Jet Propulsion Laboratory, Pasadena, California, 26 March 1979.

[8]de Groot, N. F. (JPL letter 3396-79-057; internal document), Jet Propulsion Laboratory, Pasadena, California, 20 April 1979.

[9]Bayley, W. H., Applicability of the 32 GHz Frequency Region to Deep Space Communications (JPL letter 400-WHB:jr; internal document), Jet Propulsion Laboratory, Pasadena, California, 9 July 1979.

[10]Kimball, H. G., Requirement for Deep Space Communications Allocation in the 32 GHz Region (NASA letter TN/2498; internal document), Jet Propulsion Laboratory, Pasadena, California, 17 July 1979.

[11]Bayley, W. H. (JPL letter 400-WHB:baw; internal document), Jet Propulsion Laboratory, Pasadena, California, 21 September 1979.

international treaty, ratified by most of the 160 administrations (countries) that are members of the ITU.

Assignment of radio frequencies for particular radio stations is made by national administrations, not by the ITU. These frequency assignments are expected to be in harmony with the Table of Allocations that specifies which radio services may operate in particular bands.

In addition to the international Table of Allocations, each administration has a domestic table, which may impose additional restrictions on frequency assignment within that administration.

The allocations that affect deep space research in the 32 and 34 GHz bands are particularly complex, because of variations included in the international Table and as further constrained by rules in some administrations.

In the tables that follow, radio services shown in upper case have primary status. That is, they enjoy equality with other primary services in resolving interference problems. Secondary allocations, are shown in lower case. These allocations allow operation, but give no protection with respect to primary users. Secondary users may not complain of interference, and must cease operation if they cause interference.

The numbers shown under the service allocations refer to footnotes; except where clearly irrelevant, the text of the listed footnotes is given following the statement of allocation. In a few cases, there are references to footnotes that lead down a path of further references; these paths are not followed here and should not impede appreciation of the material.

A. ITU Band Allocations

The following are ITU band allocations:

Band, GHz	Allocation
31.8–32	RADIONAVIGATION Space Research 890 891 892
32–32.3	INTER-SATELLITE RADIONAVIGATION Space Research 890 891 892 893
34.2–35.2	RADIOLOCATION Space Research 894 895 896

890   Different category of service: in Australia, Spain and the United States, the allocation of the band 31.8–32.3 GHz to the space research service (deep space) in the space–to–Earth direction is on a primary basis (see No. 425). This use shall not impose power flux-density constraints on the inter-satellite service in the band 32–32.3 GHz.

891   Different category of service: in Bulgaria, Cuba, Hungary, Mongolia, Poland, German Democratic Republic, Czechoslovakia, and U.S.S.R., the allocation of the band 31.8–32.3 GHz to the space research service is on a primary basis (see No. 425).

892   Subject to agreement obtained under the procedure set forth in Article 14, the band 31.8–33.8 GHz may also be used in Japan for space–to–Earth transmissions in the fixed-satellite service up to December 31, 1990.

893   In designing systems for the inter-satellite and radionavigation services in the band 32–33 GHz, administrations shall take all necessary measures to prevent harmful interference between these two services, bearing in mind the safety aspects of the radionavigation service (see Recommendation 707).

894   Additional allocation: in Afghanistan, Saudi Arabia, Bahrain, Bangladesh, Egypt, United Arab Emirates, Spain, Finland, Gabon, Guinea, Indonesia, Iran, Iraq, Israel, Kenya, Kuwait, the Lebanon, Libya, Malaysia, Malawi, Mali, Malta, Morocco, Mauritania, Nepal, Niger, Nigeria, Oman, Pakistan, Philippines, Qatar, Syria, Senegal, Singapore, Somalia, Sudan, Sri Lanka, Sweden, Tanzania, Thailand, Togo, Tunisia, Yemen A.R., and Zaire, the band 33.4–36 GHz is also allocated to the fixed and mobile services on a primary basis.

895   Different category of service: in Australia, Spain, and United States, the allocation of the band 34.2–34.7 GHz to the space research (deep space) (Earth–to–space) service is on a primary basis (see No. 425).

896   Different category of service: in Bulgaria, Cuba, Hungary, Poland, Mongolia, German Democratic Republic, Czechoslovakia, and U.S.S.R., the allocation of the band 34.2–35.2 GHz to the space research service is on a primary basis (see No. 425).

No. 425 is a regulatory provision dealing with footnote allocations in particular countries.

Article 14 is a regulatory provision that dictates a procedure for the coordination of frequency assignments by administrations.

B. Band Allocations by the United States

Within the U.S., bands are allocated as listed in Ref. 4. These allocations are necessarily in the context of the ITU Radio Regulations, but impose additional conditions. In some cases, there are separate allocations for government and non-government uses, as indicated below. Where no distinction is given, the allocation applies to both situations.

Band, GHz	Allocation
31.8–32	RADIONAVIGATION US69 US211 US262
32–32.3	RADIONAVIGATION INTER-SATELLITE US69 US262 US278 893
33.4–36.0	RADIOLOCATION (government) Radiolocation (non-government) US110 US252 897 G34

US69   In the band 31.8–33.4 GHz, ground-based radionavigation aids are not permitted except where they operate in co-operation with airborne or shipborne radionavigation devices.

US110   In the frequency bands ... and 33.4–36 GHz, the non-Government radiolocation service shall be secondary to the Government radiolocation service....

US211   In the bands ... 31.8–32 GHz ... applicants for airborne or space station assignments are urged to take all practicable steps to protect radio astronomy observations in the adjacent bands from harmful interference; however, US74 applies.

US252   The bands ... 34.2–34.7 GHz are also allocated for earth–to–space transmissions in the Space Research Service, limited to deep space communications at Goldstone, California.

US262   The band 31.8–32.2 GHz is also allocated for space–to–earth transmissions in the Space Research Service, limited to deep space communications at Goldstone, California.

US278   In the ... and 32–33 GHz bands, nongeostationary intersatellite links may operate on a secondary basis to geostationary satellite links.

G34   In the band 34.4–34.5 GHz, weather radars on board meteorological satellites for cloud detection are authorized to operate on the basis of equality with military radiolocation devices. All other non-military radiolocation in the band 33.4–36 GHz shall be secondary to the military services.

893   See ITU footnotes.

897   This ITU footnote does not apply to the deep space portion of the U.S. 33.4–36 GHz allocation.

C. Band Allocations by Australia

Band allocations for Australian users are given in Ref. 5.

Band, GHz	Allocation
31.8–32	RADIONAVIGATION SPACE RESEARCH (deep space) (space–to–Earth) Space research
32–32.3	INTER-SATELLITE RADIONAVIGATION SPACE RESEARCH (deep space) (space–to–Earth) AUS45 Space research
34.2–34.7	RADIOLOCATION AUS11 SPACE RESEARCH (deep space) (Earth–to–space) Space research

AUS11   Assignments to users other than the Department of Defense will not normally be authorized for this service.

AUS45   The space research (deep space) service in the space–to–Earth direction shall not impose power flux density constraints on the inner-satellite service in the band 32–32.3 GHz.

D. Allocations by Spain

Deep Space allocations by the Spanish administration are essentially those of the ITU Table of Allocations.

References

1. Hunter, J. A., *Orbiting Deep Space Relay Station Study Final Report*, Jet Propulsion Laboratory Publication 79-30, Jet Propulsion Laboratory, Pasadena, California, June 15, 1979.

2. Koerner, M. A., "Relative Performance of X-band and 32 GHz Telemetry Links on the Basis of Total Data Return Per Pass," *TDA Progress Report 42-87*, Jet Propulsion Laboratory, Pasadena, California, pp. 65-80, November–December 1986.

3. *Radio Regulations*, International Telecommunication Union, Geneva, 1982, Revised 1985.

4. *Manual of Regulations and Procedures for Federal Radio Frequency Management*, National Telecommunications and Information Administration, United States Department of Commerce, Washington, DC, 1985.

5. *Australian Table of Frequency Allocations*, Department of Communications, Australian Government Publishing Service, Canberra, Australia, 1982.

Ka-Band (32 GHz) Benefits to Planned Missions

D. M. Hansen and A. J. Kliore
Telecommunications Systems Section

This article documents the benefits of using 32 GHz downlinks for a set of deep space missions, as well as the implications to radio science and the DSN. The basic comparison is between the use of the current X-band (8.4 GHz) and a 32 GHz (Ka-band) downlink. There has been shown to be approximately an 8 dB (about 600%) link advantage for 32 GHz. This 8 dB advantage could be able to either reduce mission cost or improve mission science return.

Included here are studies on how the 8 dB advantage would be used for the Cassini and Mars Sample Return missions. While the work is preliminary, it shows that the 8 dB advantage can be exploited to provide large benefits to future deep space missions. There can be significant mass and/or power savings to the spacecraft, which can translate into a cost savings. Alternatively, the increased downlink telecommunications performance can provide a greater science return.

I. Introduction

The advantage of a higher link frequency comes from the fact that antenna gain increases in proportion to the square of the link frequency. While free space path loss also increases as frequency squared, there is a net advantage when the link employs a directive antenna at each end.

Implicit in the higher antenna gain is a narrower antenna beamwidth, which may make the task of antenna pointing more difficult. Also the effects of rain, clouds, and other atmospheric impairments are more significant at 32 GHz than at X-band. Further, antenna surfaces and structures must conform to closer tolerances to provide good performance.

The downlink frequencies of NASA deep space missions have increased from L-band (0.96 GHz) in the early Ranger days, through S-band (2.3 GHz), to the current use of X-band (8.4 GHz). Presently the technology of Ka-band (see Ref. 1 for specific frequency bands of interest) is becoming mature enough for serious consideration for near term missions. While optical frequencies may eventually provide even more advantage, optical communications technology is not expected to be ready for deep space applications for some time. The advantages of Ka-band over X-band, as well as its technological readiness appear to warrant its use in the next generation of deep space missions (Ref. 2).

Koerner (Ref. 3) has compared link performance at X-band and 32 GHz on the basis of fixed data volume during a DSN station pass, using various data rate strategies. The performance advantage of 32 GHz over X-band is very dependent on declination and DSN station location. However, over a broad

range of anticipated declinations, Koerner found at least one DSN station that provided at least an 8 dB advantage.

This article documents the analyses done in 1985 and 1986 on how the 8 dB advantage would be used for the Cassini and Mars Sample Return missions. It builds upon an earlier report by Dickinson (Ref. 4). There the benefits of 32 GHz operation were determined in terms of the minimum cost to both flight and ground systems. Here the criteria are what benefits the use of 32 GHz provides to flight projects and how various missions can use these benefits.

Sections I and II analyze the use of the 8 dB performance advantage for the Cassini and Mars Sample Return (MSR) missions, respectively. For Cassini, the use of 32 GHz allows for an increase in the data rate by a factor of five, if the baseline high gain antenna (HGA) is retained along with the baseline dc power allocation for communications. Alternatively the performance advantage of a 32 GHz link allows the use of a smaller antenna. The most attractive option is to reduce the output RF power (lower DC power), which provides a net cost savings to the spacecraft after the nonrecurring costs are paid because a 25 W RTG[1] unit can be removed.

For MSR, landed mass and size (for packaging considerations) are extremely important drivers. Under the current mission scenario the rover will only communicate with the ground when it is not moving; hence the communications hardware can use the power allotted for the locomotion function, so DC power is not a spacecraft driver. The 32 GHz frequency allows for the use of a smaller, lighter antenna than with X-band. Two options are studied. One uses a parabolic reflector; the second uses a flat plate array. The array provides the better mass and size performance of the two.

Section III analyzes the benefits of 32 GHz to radio science. It discusses mission dependent benefits for gravity wave and relativity experiments, solar corona studies and bistatic radar. While 32 GHz is not beneficial for all types of radio science, for the experiments listed above, the higher frequency reduces solar plasma and atmospheric effects on the signal. Section IV illustrates the relief available in DSN loading at 32 GHz in comparison to 8.4 GHz.

II. The 32 GHz Benefits for Cassini

The baseline Cassini telecommunications system design has an 8.4 GHz high rate downlink that uses the Voyager HGA

(3.7-meter) and redundant 10.6 W X-band solid-state power amplifiers (XSSPA).

The spacecraft effective isotropic radiated power (EIRP) at X-band is 87.73 dBm (see Table 1). The antenna aperture efficiency is 72.5%, based on measured data. The pointing loss assumes upgraded sensors from the Comet Rendezvous Asteroid Flyby (CRAF) mission baseline to achieve a boresight error of 0.134 degrees. This EIRP provides a nominal downlink data rate of 30 kbps at X-band.

Based on Koerner (Ref. 3), an 8 dB advantage is assumed. The Cassini mission will be largely near a +20 degree declination. At this declination the northern stations will have a 9 dB 32/8.4 GHz advantage with 90% link confidence whereas Canberra will have about a 7 dB advantage. Thus, 8 dB is a good approximation for illustrative purposes. Other assumptions for the 32 GHz design are as follows: (a) the 32 GHz power conversion efficiency is 21%; (b) the 32 GHz exciter power is the same as the 8.4 GHz exciter power; (c) Voyager HGA efficiency is the same at 32 GHz as it is for 8.4 GHz (Cassegrain feed); and (d) a 2.3 GHz downlink is transmitted via the HGA for radio science.

Table 2 shows a number of options for the telecommunication subsystem on the Cassini spacecraft using various combinations of 32 and 8.4 GHz equipment.

A. Option #1: Current Cassini Baseline

The baseline for Cassini is the all X-band configuration as described earlier.

B. Option #2: All 32 GHz System

The X-band downlink is deleted entirely. An array feed power amplifier (AFPA) with electronic beam steering (EBS) is assumed. The AFPA has 21 elements.[2] The EBS requires that the spacecraft have a fiber optic rotation sensor (FORS) to provide precise pointing knowledge. The antenna pointing calibrations will be better due to the narrower beam. This leads to a pointing error of 0.107 degree and, with the EBS, a pointing loss of 0.50 dB.[2]

It is assumed that the antenna aperture efficiency is the same at 32 GHz as at 8.4 GHz so the 32/8.4 GHz advantage is 8.0 dB. There are no circuit losses for the AFPA, so to achieve

---

[1] Radio isotope thermonuclear generator (RTG) is a source of onboard dc power.

[2] Boreham, J. F., "A 21 Element EBS Array Feed for the SOTP Spacecraft," JPL IOM 3360-85-030 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 22, 1985, and Boreham, J. F., "A Further Explanation of Ka-band Spacecraft HGA Pointing Control Options," JPL IOM 3360-85-033 (internal document), Jet Propulsion Laboratory, Pasadena, California, December 10, 1985.

the same received SNR on the ground the spacecraft RF power can be reduced by 8.82 dB (includes circuit losses and the difference in pointing error).

Boreham[2] has suggested that an allowance for one module failure be made. This increases the transmitted power by 0.44 dB. This means that the 32 GHz power level is 8.38 dB less than at X-band or 1.5 W. Assuming a 21% power conversion efficiency this requires 7.3 W of DC power. The transmitter mass is 3.0 kg. There is an increase in the mass (+2 kg) and power (+2 W) for the EBS.

Deletion of the redundant 10.6 W XSSPAs results in a mass savings of 5.4 kg. Overall, this results in a small mass savings for the radio frequency subsystem (RFS). The required DC power drops from 40 W to 9.3 W for a savings of 30 W. This allows for dropping a 25 W RTG. Priced at $200K/W this means a $5M savings.

The nonrecurring cost for the 32 GHz hardware is $7M. The recurring cost is $4.4M. There is a savings of $1M for dropping the 10.6 W XSSPAs. The mass, power and cost deltas are listed in Table 2.

C. Option #3: 32 GHz Prime With Minimum 8.4 GHz Backup

While replacing the 8.4 GHz downlink with an all 32 GHz system appears reasonable there are good reasons to keep an X-band system aboard. The X-band low gain antennas (LGAs) are on-board the spacecraft for the uplink. The 32 GHz LGAs for a near-earth downlink would have to be added. The performance of an LGA 32 GHz link may be 3–4 dB worse than a comparable X-band LGA link. This may require use of the HGA more during the near-earth phase.

A 32 GHz redundant link through an LGA may require separate amplifiers. The AFPA cannot be used for a 32 GHz LGA link. There could also be large circuit losses for the cable runs to the LGAs. If 32 GHz TWTAs were used with the HGA they could also be used for the LGAs, but then a different type of pointing system would be required. Pulse plasma thrusters or reaction wheels would provide the necessary pointing precision but their mass and/or power penalties are very severe. Thus, to the AFPA/EBS system considered in Option #2 is added an X-band capability. Specifically, 3.0 W XSSPAs are added to Option #2 for the near-earth link and as a backup for the 32 GHz downlink. These are modules in the CRAF 5.6 W XSSPA. The X-band RF power is decreased by 5.5 dB. There is also a 1 dB loss in X-band gain for the HGA because the feed is switched from Cassegrain to focal point.

It is estimated that the 3.0 W XSSPAs will add 1.5 kg of mass and that each will consume 10 W. The nonrecurring and

recurring costs are $0.5M and $0.4M respectively.[3] There is only a slight increase in required power (9.3 W – 10 W) because the EBS is not needed for X-band operation.

D. Option #4: 32 GHz Prime With Higher Power 8.4 GHz Backup

The baseline CRAF 5.6 W amplifiers are used instead of the lower power 3.0 W XSSPAs of Option #3. No additional non-recurring costs should be required. The recurring cost for a pair of 5.6 W XSSPAs is $600K. The 5.6 W XSSPA requires 20 W of DC power, which is 10 W more than the 32 GHz AFPA/EBS. Power sharing on the spacecraft is required to use this amplifier as a backup. By adding a switch and an ortho-mode feed the power from both XSSPAs could be summed to provide near baseline X-band perfomance. This will, of course, require more DC power. The two 5.6 W XSSPAs add 3.6 kg of mass.

E. Option #5: 32 GHz Prime With Half-Size Antenna

A smaller (1.7-meter) antenna is less expensive than the 3.7-meter size, and permits more relaxed pointing; EBS is not required. The baseline AACS system is adequate. It is assumed that the baseline 10.6 W XSSPAs are kept onboard the spacecraft.

The 32 GHz RF power is determined by assuming use of the 40 W DC prime power required for the X-band baseline and the 21% power conversion efficiency. This gives an RF power of 8.4 W. At this power level a 1.7-meter dish is adequate to achieve the same received SNR. This uses an AFPA with no circuit loss and an antenna with aperture efficiency of 60%.

According to Dickinson (Ref. 4), the reduction in antenna size will save 18 kg and $600K. The non-EBS AFPA nonrecurring and recurring costs are $5.8M and $3.4M, respectively. The mass is 3.0 kg. These numbers are all for a 21-element array. However, the higher power level may require more than 21 elements in the array. Hence both the costs and mass estimates may be higher.

The smaller antenna reduces the X-band and S-band HGA performance by about 7 dB. It may also allow for a different spacecraft design using a gimballed antenna instead of a body-fixed Voyager HGA. Figures 1 and 2 show the baseline Cassini spacecraft and the same spacecraft with a smaller antenna, respectively.

[3]Personal communication with A. L. Riley, Spacecraft Telecommunications Equipment Section, December 1985.

F. Option #6: 32 GHz Performance Augmentation

This option adds a full capability 32 GHz system to the baseline Option #1. This is the most expensive option. By using all of the available DC power and subtracting 2 W for the EBS, 8 W at 32 GHz could be generated. This allows for returning the baseline data rate (30 kbps) into a 34-meter station or increasing the data rate by a factor of five.

The X-band performance is reduced by about 1 dB by switching to a focal point feed, a move necessitated by the 32 GHz AFPA at the cassegrainian focus.

As in option 5, the nonrecurring and recurring costs may be greater for this higher power option. More elements will probably be needed because of the higher power requirement.

III. The 32 GHz Benefits for the Mars Sample Return Mission

A. Introduction

This section compares the mass and DC power requirements of two telemetry system designs for a Mars Sample Return Rover. The two frequencies are the current 8.4 GHz and the proposed 32 GHz. Comparisons are made also for the 34-meter high efficiency DSN antenna and the 70-meter DSN antenna subnets.

The Rover mission is conceived as highly interactive with the Earth-based operations. Images of a possible route are sent to Earth. Based on decisions there, commands are sent to the rover. It moves, stops, takes pictures and starts the cycle over again. Movements are limited to the Mars horizon, approximately 0.245 km, 10 minutes for playback per hour, Mars daytime only, and no operation in potential dust storms. These characteristics together with credible imaging system properties at 120 kbps and a 4:1 data compression algorithm yield a Mars-to-Earth channel operating rate of 30 kbps.

A key constraint is the rover antenna envelope which must be less than 1.4-meter in maximum dimension (diagonal/diameter) because of packaging considerations. The last key assumption is that the entire locomotion power of 120 W raw DC is available for downlink communications when the Rover is stopped.

Two types of rover-borne transmission options are evaluated. These are a parabolic antenna with traveling wave tube amplifier attached, and a flat planar array antenna. The designs are evaluated at both frequencies and for reception by both 34-meter HEF and 70-meter DSN stations.

B. Methodology

Telecommunications link design control tables yield transmitter power gain products required to communicate from the Rover to each DSN aperture size. Then, by using mass relationships for antenna size, TWTA power level, heat radiation, structure and DC power, system mass is minimized subject to the constraint that DC power be less than or equal to 120 W and that the envelope be less than or equal to 1.4-meter diagonal/diameter.

Figures 3 and 4 show the mass minimization results for each frequency and DSN antenna combination with each Rover antenna type, respectively. Figures 5 and 6 show the spacecraft antenna aperture area results corresponding to the minimum mass configurations. Figures 7 and 8 show the DC power results corresponding to the minimum mass configurations. In all cases, the DC power used is less than the 120 watts allowed.

This analysis shows a clear mass advantage of the 32 GHz system over the 8.4 GHz system with either antenna type. This preference is increased when communication through the 34-meter apertures is required. Other advantages attendant to 32 GHz design are that (1) mission reliability will improve if both DSN aperture types can support the mission; (2) a 120 kbps high activity science mode can be supported by the 70-meter aperture; and (3) network loading in the 2000 AD era, although not a demonstrable problem now, would favor being able to operate with 34-meter capability.

Between the two antenna types, other packaging considerations favor the 32 GHz array. The entry aeroshell imposes a volume constraint as well as an area constraint. A flat plate array occupies less volume than a parabolic antenna of the same area. Furthermore, the area constraint becomes important at 0.8 m^2 and absolute at 1.6 m^2. In all cases 32 GHz array is less than or equal to 0.4 m^2.

IV. The 32 GHz Benefits for Radio Science

The term radio science encompasses a number of diverse disciplines that have distinctly different and often conflicting requirements. Therefore, the magnitude of improvement that can be expected from the introduction of a 32 GHz downlink depends on the specifics of individual experiments.

For example, the very effects of propagation through refractive plasma regions that are the objectives of those who wish to study the solar wind and interplanetary plasma are regarded as corrupting noise by those who wish to use the radio link to search for gravitational waves and to test competing relativity theories. By virtue of its higher frequency, 32 GHz

offers a reduction in noise caused by interplanetary plasma, but at the same time it is less sensitive to the effects of tenuous plasma regions that are desirable targets of observation by propagation scientists, such as planetary nightside ionospheres and cometary ion tails. However, in order to accurately measure such tenuous plasma regions, it is necessary to use two coherent downlink frequencies, and it will be the other (lower) frequency such as S-band, for example, which will ultimately determine the sensitivity to tenuous plasma.

In the following paragraphs, each major area of radio science is reviewed relative to the impact of a 32 GHz downlink.

A. Gravitational Waves and Relativity

In this area, the reduction of noise due to the interplanetary plasma would improve the detectability of gravitational waves. Note that the improvement would be far greater if a Ka-band uplink could also be used.

B. Solar Corona Studies

For these measurements, which are concerned with the structure and turbulence in the solar corona, 32 GHz would allow deeper penetration. (The present limit with X-band is about 1.4 solar radii.)

C. Gravity

The presence of 32 GHz on the downlink would help reduce the noise on the Doppler data that are used to estimate the masses of planetary, asteroidal, and cometary bodies, and hence lead to more accurate estimates.

D. Propagation Science

In this area, which deals with planetary and cometary atmospheres and ionospheres through their effect on the propagation of spacecraft signals, the impact of 32 GHz is roughly neutral. In the case of planetary atmospheres, the greater gain of similarly sized spacecraft antennas at 32 GHz would increase the available signal margin, but for deep penetration in atmospheres such as those of Venus and the outer planets, much of this margin would be cancelled because of increased losses due to absorption and scattering. In the area of measuring tenuous dispersive media such as planetary and cometary ionospheres, a lower frequency (such as S-band or L-band) is required in addition to the 32 GHz and it is this lower frequency that determines the sensitivity.

E. Bistatic Radar

The roughness characteristics of planetary, cometary and asteroidal surfaces can be investigated by means of bistatic radar. For close encounters, such as cometary rendezvous missions, the narrow beamwidth of a 32 GHz downlink would lead to improved spatial resolution.

V. The 32 GHz Benefits for Network Loading

The attractiveness of 32 GHz for downlink channels is illustrated in Fig. 9. It shows that at 32 GHz, a 70-meter station has the equivalent aperture of an array at X-band of a 70-meter, a 32-meter high efficiency (HEF), and, a 34-meter (all at Goldstone) plus the array of antennas at the Very Large Array (VLA) in New Mexico. This array will be used for the Voyager Neptune encounter. It is a special one-time event. A 70-meter at 32 GHz would provide the same capability on a continuous basis for future deep space missions.

Also, at 32 GHz a 34-meter HEF is equivalent to an X-band 70-meter antenna in terms of receiving capability. This is also shown in Fig. 9. The use of a 34-meter antenna reduces commitments for 70-meter support. It provides missions and the DSN some flexibility in providing necessary mission support, and offers potential for relieving network loading conflicts.

VI. Conclusion

This article has looked at using 32 GHz as the downlink frequency for future deep space missions, utilizing the 8 dB performance advantage for 32 GHz relative to 8.4 GHz.

The two missions discussed, Cassini and MSR, could readily make use of this 8 dB. Cassini would probably use a lower transmit power that would allow for dropping an RTG from the spacecraft. Alternatively, the size of the antenna could be reduced, which would permit a different spacecraft configuration. For MSR, power is not a driver; mass is. The use of 32 GHz would allow for a smaller, less massive antenna to be flown.

This article has documented work done on 32 GHz in 1985 and 1986. In the next few years, better cost and performance estimates will become available with more analysis as the Ka-band development evolves.

References

1. de Groot, N. F., "Ka-band (32 GHz) Allocations for Deep Space," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

2. Smith, J. G., "Ka-band (32 GHz) Downlink Capability for Deep Space Communications," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

3. Koerner, M. A., "Relative Performance of X-band and 32 GHz Telemetry Links on the Basis of Total Data Return Per Pass," *TDA Progress Report 42-87*, Jet Propulsion Laboratory, Pasadena, California, pp. 65–80, November 1986.

4. Dickinson, R. M., "Comparison of 8.415-, 32.0- and 565646-GHz Deep Space Telemetry Links," *JPL Publication 85-71*, Jet Propulsion Laboratory, Pasadena, California, October 15, 1985.

Table 1. Cassini EIRP

Parameter	Design Values	Tolerances
10.6 W XSSPA	40.25 dBm	±1.00 dB
Transmit CKT Loss	−0.45 dB	±0.20 dB
Antenna CKT Loss	−0.30 dB	±0.10 dB
Antenna Gain	48.80 dBi	±0.50 dB
Antenna Pointing Loss	−0.57 dB	(0.134° of boresight, maximum value)
Total	87.73 dBm	

Table 2. Cassini 32-GHz options

Option	Antenna Size, m	Frequency, GHz	Antenna Pointing, deg/Loss, dB	RF Power, W	DC Power, W	Transmitter Mass, kg	AACS Mass, kg	AACS Power, W	Δ Mass, kg	Δ Power, W	Δ Cost* Power	Hardware Non-recurring	Recurring
(1) X-Band Baseline	3.7	8.415	0.134/−0.57	10.6	40	5.4	10	10	−−	−−	−−	−−	−−
(2) 32 GHz with AFPA/EBS	3.7	32	0.107/−0.50	1.5	7.3	3.0	12	12	−0.4	−30.7	($5M)	$7M	$3.4M
(3) Add 3.0 W XSPPAs to #2	3.7	8.415	0.107/−0.36	3.0	10	4.5	12	10	+1.1	−30.7	($5M)	$7.5M	$3.8M†
(4) Add 5.6 W XSSPAs to #2	3.7	8.415	0.107/−0.36	5.6	20	6.6	12	10	+3.2	−30.7	($5M)	$7M	$4.0M†
(5) Smaller Antenna 10.6 W XSSPAs	1.7	32	0.107/−0.96	8.4	40	8.4	10	10	−15.0	0	0	$5.8M	$2.8M
(6) Add Ka-Band AFPA to Baseline	3.7	32	0.107/−0.50	8.0	38	8.4	12	12	+5	0	0	$7M	$4.4M†

*Tolerances on cost numbers are +30% and −10%.

†Dual frequency S/X feed, TBD dollars.
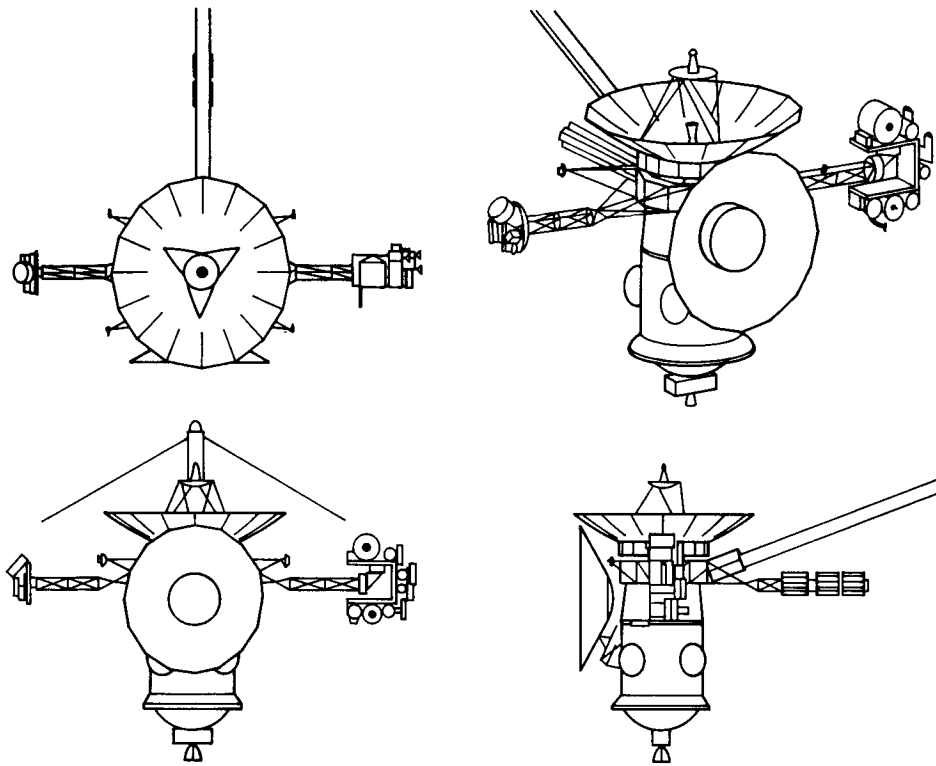
Fig. 1. Cassini spacecraft with 3.66 m HGA
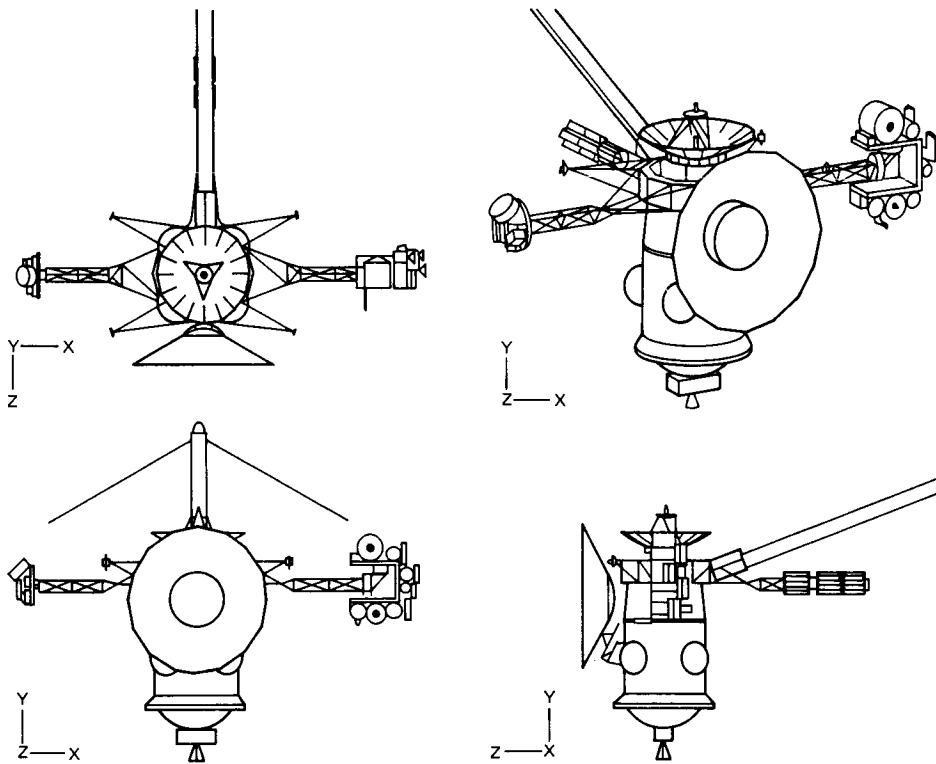
Fig. 2. Cassini spacecraft with 1.7 m HGA
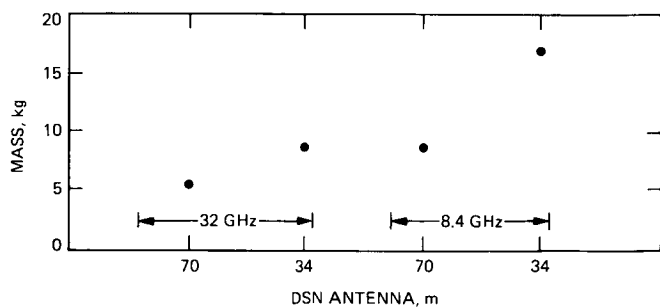
Fig. 3. Minimum mass of parabolic antenna with TWTA

Fig. 4. Minimum mass of flat planar array
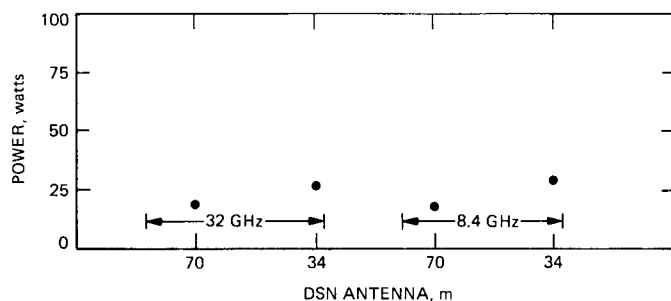
Fig. 5. Area of minimum mass configuration for parabolic antenna with TWTA

Fig. 6. Area of minimum mass configuration for flat plate array

Fig. 7. DC power for minimum mass configuration for parabolic antenna with TWTA
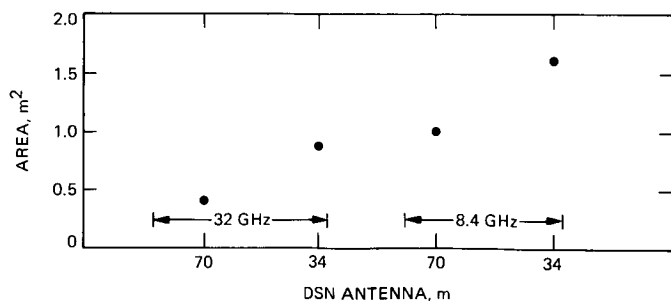
Fig. 8. DC power for minimum mass configuration for flat plate array

118
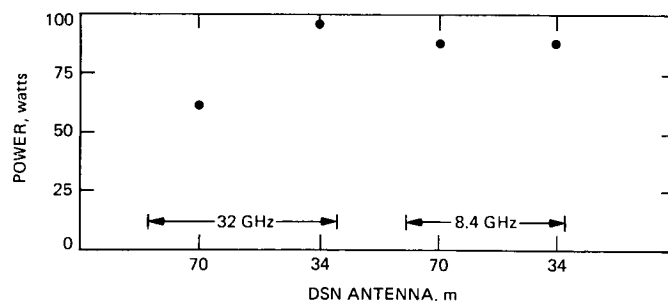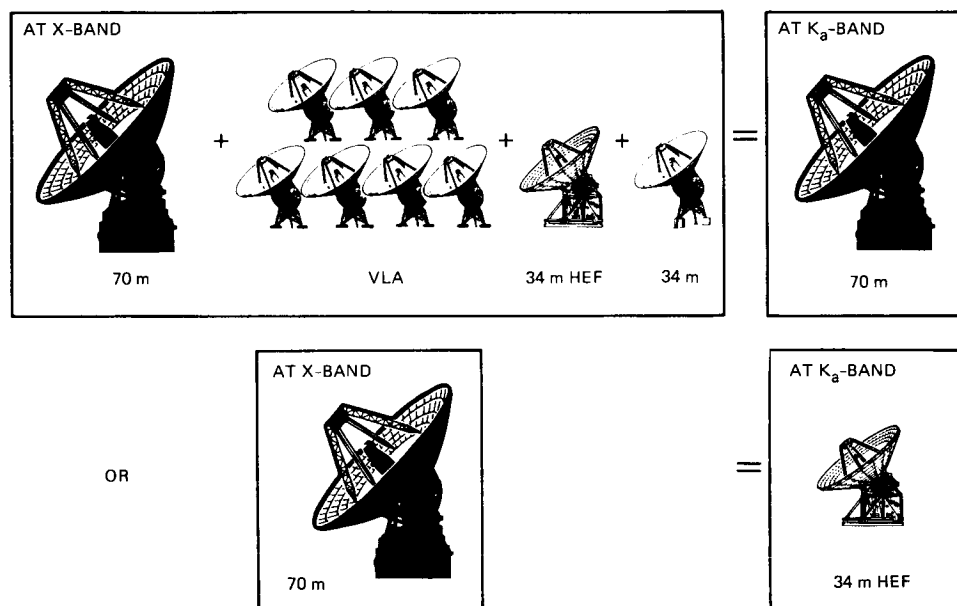
Fig. 9. The G/T equivalence for X-band and 32 GHz

A Growth Path for Deep Space Communications

J. W. Layland
TDA Planning Office

J. G. Smith
TDA Technology Development Office

Increased DSN receiving capability far beyond that now available for Voyager is achievable through a mix of increased antenna aperture and increased frequency of operation. In this note we consider a sequence of options: (1) adding mid-sized antennas for arraying with the existing network at X-band; (2) converting to Ka-band and adding array elements; (3) augmenting the DSN with an orbiting Ka-band station; and (4) augmenting the DSN with an optical receiving capability, either on the ground or in space.

Costs of these options are compared as means of achieving significantly increased receiving capability. The envelope of lowest costs projects a possible path for moving from X-band to Ka-band and thence to optical frequencies, and potentially for moving from ground-based to space-based apertures. The move to Ka-band is clearly of value now, with development of optical communications technology a good investment for the future.

I. Introduction

The Deep Space Network (DSN) configuration which supported the encounter of Voyager with Uranus is the most sensitive and capable receiving instrument for deep space communications assembled to date. Even so, more data could have been gathered by the Voyager's instruments had a more capable link been available to transport that data to Earth. Other instruments are available now that could (if permitted) drive the communications link toward significantly increased capability. Is there a good technology pathway to follow to achieve significant growth in capability? We examine that question in the following and conclude that indeed, the development of Ka-band now, and research work leading toward optical/visible-band communications for the not too distant future, are both elements of that pathway. The material presented here is not brand new, having been adapted from results available in our references, but it is examined from a substantially different perspective.

The top level characteristics of the DSN and its relationship to its flight mission customers have developed along a well established philosophy of steady dependable service and continuous evolutionary growth of capabilities (Long Range Plan for the Deep Space Network (JPL Internal Document), Jet Propulsion Laboratory, Pasadena, California, October 1984). That philosophy is accommodated here, by building

upon the existing infrastructure, including DSN capabilities and DSN and flight mission operating strategies. The DSN of 1986 provides support to spacecraft operating at X-band (8 GHz) and S-band (2 GHz) from three communication complexes spaced approximately equally in longitude about the Earth so that continuous coverage can be available if spacecraft events justify it. Each of the three complexes provides essentially the same capability as any other, so that the scheduling of spacecraft events is dependent only upon specifics of the trajectory and not upon network constraints. The parameters of the communication link are chosen such that there is at all times a very high confidence (90-95% or more) in achieving adequate communication. Variational tolerances of equipment performance and weather-dependent effects are included here.

Improvements to DSN capabilities have been incremental, built upon the existing features and adding new ones as needed for the next flight mission. Old features get discarded only when old in-flight missions dependent upon them cease to function. New features are examined and justified for their value to forthcoming missions, as compared to costs for other ways of achieving comparable mission objectives. For the purposes of this report, we consider the incremental cost to NASA of providing a significantly increased communications link capability for the "next" mission by means of a few tightly constrained pathways. Features of the 1986 DSN and of the current-design spacecraft are assumed to be available without added cost. High-confidence 24-hour coverage is assumed to be required.

II. Options for Growth

Options for communication growth examined here are of two types—increase the ground receiving aperture, or increase the communication frequency band used. Other parameters, such as spacecraft power, transmitting aperture, etc., are held constant at values which are presumed reasonable. The frequency bands of interest are the same as those examined by Dickinson (Ref. 1), who compared costs of X-band, Ka-band, and Optical communication for a fixed total data volume. According to the study done in the late 1970's of the Large Advanced Antenna Station (LAAS) (Haglund, H. H., et al., Large Advanced Antenna Station Status Report (JPL Internal Document No. 890-74), Jet Propulsion Laboratory, Pasadena, California, August, 1978), large receiving apertures at X-band are best achieved via arraying of moderate sized antennas of 30- to 40-meter diameter. The effective aperture is a function of the number of antennas employed. The successes at Voyager encounter demonstrated the utility of arraying in fact, as well as theory. We presume the same will hold at the higher frequency regimes as well.

Increasing the frequency band of operation improves the communication performance by narrowing the transmitted beam, thus delivering a larger fraction of the power to the receiving aperture, assuming all else is fixed. This requires a concomitant improvement in the precision with which the transmitted beam is pointed at Earth, thus imposing a requirement on spacecraft design that we cannot really deal with here. Thus it is assumed for this examination that the spacecraft attitude control is retained precisely enough for the body-fixed X-band (8.4 GHz) transmission, and that greater precision as needed is provided by the communication subsystem itself. For Ka-band (32 GHz), the increased precision of pointing is a factor of four as compared to X-band. An array feed solid state power amplifier with electronic beam steering, or a beam waveguide mirror system, is believed capable of this pointing refinement if it is provided with adequate knowledge of the true spacecraft attitude. The pointing of the optical (560 THz) transmission beam must be more than three orders of magnitude more precise than that for X-band. It has frequently been argued that the required pointing can be accomplished, again by beam waveguides, i.e., by steering the mirrors of the transmitting optical telescope, which are now much smaller and lighter than the comparable microwave components. We presume this to be the case, but retain a concern which will appear in the uncertainties of the cost-performance curves to be displayed.

The baseline capability against which other options are to be compared is that of the 1985-86 vintage DSN 64-meter antenna, operating at X-band as it did to support Voyager. The reference spacecraft transmits at X-band via a 10-watt Traveling Wave Tube Amplifier (TWTA), and a 4.5-meter aperture. This is the same spacecraft configuration as that selected as optimum by Dickinson, and the antenna represents the largest non-furlable aperture which can be carried in the Shuttle payload bay. Assuming TWTA efficiency of about 40%, the raw spacecraft power into the transmitter is about 25 watts. This combination supports a 26 kbps data rate from Saturn, which is available 24 hours per day via the three complexes of the DSN.

III. Option Cost Comparisons

Individual cost and performance estimates which make up the curves (Fig. 1) of cost vs communication growth are taken largely from Dickinson (Ref. 1). For these curves, the fixed parameter is raw spacecraft power into the transmitter, which is approximately 25 watts for both the X- and Ka-band cases in the reference report, and also set here at 25 watts input to the optical transmitter with assumed 8% efficiency. The spacecraft transmitting apertures for these curves are the same as those in the reference study: 4.5-meter antenna for microwave

transmission and the 28 cm OPTRANSPAC telescope for optical. The parameters varied are the receiving aperture and the frequency band.

A. X-Band Curve

The zero-point for all curves is the capability of the DSN's current 64 meter antenna in conjunction with the referent X-band spacecraft. These antennas are currently being upgraded to become high efficiency 70-meter antennas, thus adding 2 dB in X-band communications performance. This upgrade will be completed by 1988, at a basic cost of $30M (±10%) (McClure, D. H. (JPL Private Communication) also Stevens, R., FY85 C of F Antenna Projects (JPL Presentation to NASA-OSTDS) Jet Propulsion Laboratory, Pasadena California, May 16, 1985). Further increases in X-band capability are available through arraying with an arbitrarily large number of 30- to 40-meter antennas a la the LAAS study results (Haglund, H. H., et al. Large Advanced Antenna Station Status Report (JPL Internal Document #890-74), Jet Propulsion Laboratory, Pasadena, California, August 1978). The aperture efficiency and system temperature of these antennas would be the same as that of the upgraded large antennas. The 40-meter antennas at last look were priced at $10M (assume ±10%) each, including all electronics needed for operation and arraying (Stevens, R., Report: Use of VLA and Japanese 64m vs Temporary New DSN Implementation for Voyager Neptune Encounter Support (JPL Internal Document IOM #RS84-7051D), Jet Propulsion Laboratory, Pasadena, California, October 29, 1984). This figure is consistent with the engineering estimates contained within the LAAS study, adjusted as necessary for the intervening cost inflation. Approximately half of this figure represents the steel and concrete of the radio telescope itself, while the other half is the electronics, control, and support equipment necessary to make the antenna a functioning entity. Expanding the three complexes symmetrically results in the X-band curve as shown.

B. Ka-Band Curve

The Ka-band version of the referent spacecraft uses a 5 W array feed power amplifier with the 4.5 m antenna. Raw spacecraft DC power is again approximately 25 W for the expected efficiency of this amplifier. Operation of this spacecraft communications link into the 70-meter provides a data rate of 117 kbps, when the 70-meter has been enhanced for Ka-band operation. Link margin for this rate was set to provide 90% confidence of successful communication. Cost of this capability is the $30M for the basic 70-meter upgrade, plus $25M (±20%) for their Ka enhancement, plus $59M (±$12M) for the first Ka-band capable spacecraft (Ref. 1). With this completed, the 70-meter is expected to be approximately 55% efficient at 32 GHz. Increments to this capability are assumed

to be available in the form of 40-meter antennas which are 70% efficient at a cost of $12.5M (±20%) each, or for a 25% surcharge over their X-band counterparts. With these characteristics, each 40-meter aperture adds 48 kbps to the communications capability. The Ka-band curve shows the cost-performance path for symmetric growth of the three DSN complexes at Ka-band via these 40-meter arrayable modules.

C. Space-Based Ka-Band Point

The triangle indicated as H79 shows the approximate performance and cost of the 28-meter Ka-band Orbiting Deep Space Relay Station (ODSRS) as derived from Hunter (Ref. 2). That receiving system had a G/T performance which was 6 dB above that of the 64-meter X-band capability. Overall link performance is indicated on Fig. 1 at 3 dB to account for the 3 dB lower efficiency, and hence lower output from the transmitter of the Ka-band spacecraft. The total cost of the ODSRS as perceived in 1979 was $400M, including design, implementation, launch and on-orbit assembly, and 10 years M&O. Of that figure, $120M was supposed to include three Shuttle launches, plus the Orbital Transfer Vehicle needed to place the ODSRS at Geosynchronous altitude. Shuttle launch costs are currently carried as $140M per full cargo bay (Ref. 1), while the upper stage itself should cost on the order of $60M, consistent with the now-defunct STS-Centaur (Ref. 1). Thus an updating of the launch/installation cost entry would raise it from $120M ('79) to $480M ('86). The remaining cost elements, totaling $280M ('79), are items subject to general price inflation which is a net 55% over these years, for an adjusted cost of $440M ('86), with big uncertainty. Total ODSRS cost would be on the order of $920M (±20%) in 1986. The cost indicated on Fig. 1 includes the ODSRS plus $59M for the first Ka-band user spacecraft.

D. Ground-Based Optical Curve

Ground-based optical is evaluated assuming a user spacecraft with a 28 cm transmitting telescope and a 2 watt laser transmitter, consistent with raw spacecraft power of 25 watts and efficiency of 8%. Lasers currently exist with this efficiency, but at lower power levels (Ref. 3). The unit receiving aperture is assumed to be 10-meters, the same as the Keck Telescope to be built in Hawaii, but of substantially lower optical imaging quality. To counteract cloud blockage, the receiving apertures must be at least triplicated and spatially diversified in each longitude. (An alternative is added on-board storage and time diversity.) It is believed that in quantity, these 10-meter photon buckets would cost much less than the reported $70M cost of the Keck Telescope. Total capability cost for any specific level is the sum of the first-user spacecraft cost of $88M (±$16M), plus the ground receiving system costs for the collecting aperture. Lesh (Ref. 4) has recently calculated performance of this spacecraft-ground combination

for both daylight and dark sky conditions while examining cost tradeoffs of size and surface quality for the ground receiving telescope element. To meet our requirements for continuous availability, link design for daylight conditions is appropriate. According to these calculations, a 10-meter collecting aperture with a surface quality adequate to provide 16 dB communications growth over the reference X-band system could be acquired for unit cost of $25M. To that figure we should add an overhead of about $5M for facilities, utilities, and interfacing into the remainder of the DSN's machinery, which is assumed to be available to support the optical receiving telescope as it does the radio ones today. Because of the analogy approach to these costs, an uncertainty of 2 dB (+60%/-40%) seems appropriate here. Total nominal cost for one optical subnet enabling a 16 dB communications growth thus is $270M, for three-point diversity in each of the three longitudes of the Network. Arraying these apertures provides the growth path shown. Smaller or lower quality telescopes can be used at somewhat lesser cost, as indicated by the dashed segment of the optical curve in Fig. 1.

E. Space-Based Optical Points

The orbiting optical DSS, as portrayed in early presentations of the 1985 study (Dickinson, R. M., Review of Ka-Band Study Task Results (JPL Presentation), Jet Propulsion Laboratory, Pasadena, California, March 22, 1985), can provide 761 kbps from a 130 mw laser with 28 cm transmitting telescope into an orbiting 20-meter photon bucket (LDR-type). Assuming that the transmitting laser power can be scaled upward to 2 watts, consistent with raw spacecraft power of 25 watts and efficiency of 8%, the overall communications capability becomes 26.5 dB above the X-band reference. Estimated cost of this orbiting receiving system was $300M, including $140M for the single Shuttle launch, on-orbit assembly, and installation on a space station (Dickinson, R. M., Review of Ka-Band Study Task Results (JPL Presentation), Jet Propulsion Laboratory, Pasadena, California, March 22, 1985). We have added another $140M to the installation costs as a rough estimate for the cost of placing it on a Geostationary platform instead of into low orbit. The first user spacecraft cost of $88M (±$16M) for development and integration of the optical transmitting subsystem is also included in the $530M cost denoted as OP85 on Fig. 1. Indicated uncertainty in cost is about 2 dB (+60%/-40%). As compared to the Ka-band ODSRS, the significantly lower cost is due in part to the smaller size, and in part to the

assumed existence of a fully functioning Geosynchronous Space Platform which will provide real-estate and utilities to the optical DSS.

The orbiting optical DSS as finally described by Dickinson (Ref. 1) consisted of a 4.5-meter diameter photon bucket, which would also be Space-Station mounted. This reduction in size eliminated most of the on-orbit assembly work and cut launch charges to one-third of a Shuttle bay. It also lowered the communications performance by 13 dB by virtue of the reduced collecting area. Estimated cost of this device was $145M (+$62M/-$40M), as installed on a space station in low Earth orbit. To achieve full-time coverage for a using spacecraft in deep space, we must either assume the existence of a second station in low orbit and half rotation away, or assume the existence of another station at Geosynchronous altitude and allocate a premium for transporting the optical receiver to the higher location. We have chosen the latter path, and have added another $100M to the installation costs for this purpose. The first user spacecraft cost of $88M (±$16M) is also included in the $330M (+60%/-40%) cost point denoted as OP85' on Fig. 1.

Taking Fig. 1 in its entirety, the envelope of lowest costs follows a path from X-band to Ka-band with modest levels of arraying, and thence to optical frequencies. Space-based elements, unrealistic as free-flyers for microwave frequencies, also appear of value for optical frequencies with the assumed economies of residing on an established Geosynchronous Space Station Platform.

IV. Concluding Remarks

With time, it is expected that both the Ka-band and optical transmitter efficiencies will improve, thus moving these curves to the right, and perhaps lowering their costs at the same time. The X-band curve should be reasonably stable. The fuzziest thing on these curves is the ground-based optical, with both performance and cost very uncertain. Concern exists as well in the ability to accurately and stably point the very narrow optical beam. Both Ka-band and optical pathways show significant promise for future growth in communications capability. The technology for Ka-band is almost in hand, and it should be pursued vigorously to exploit that promise. Optical communication technology makes an excellent investment for a slightly more distant future.

References

1. Dickinson, R. M., *A Comparison of 8.415-, 32.0-, and 565646- GHz Deep Space Telemetry Links*, JPL Publication 85-71, Jet Propulsion Laboratory, Pasadena, Calif., Oct. 15, 1985.

2. Hunter, J. A., et al., *Orbiting Deep Space Relay Station Study Final Report*, JPL Publication 79-30, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1979.

3. Sipes, D. L., "8.5%-Efficiency Nd:YAG Laser Development," *JPL Highlights 1985*, JPL #400-282, Jet Propulsion Laboratory, Pasadena, Calif., November, 1985.

4. Lesh, J. R., and Robinson, D. L., "A Cost-Performance Model for Ground-Based Optical Communications Receiving Telescopes," in *TDA Progress Report 42-87*, Jet Propulsion Laboratory, Pasadena, Calif., November 15, 1986, pp. 56-64.
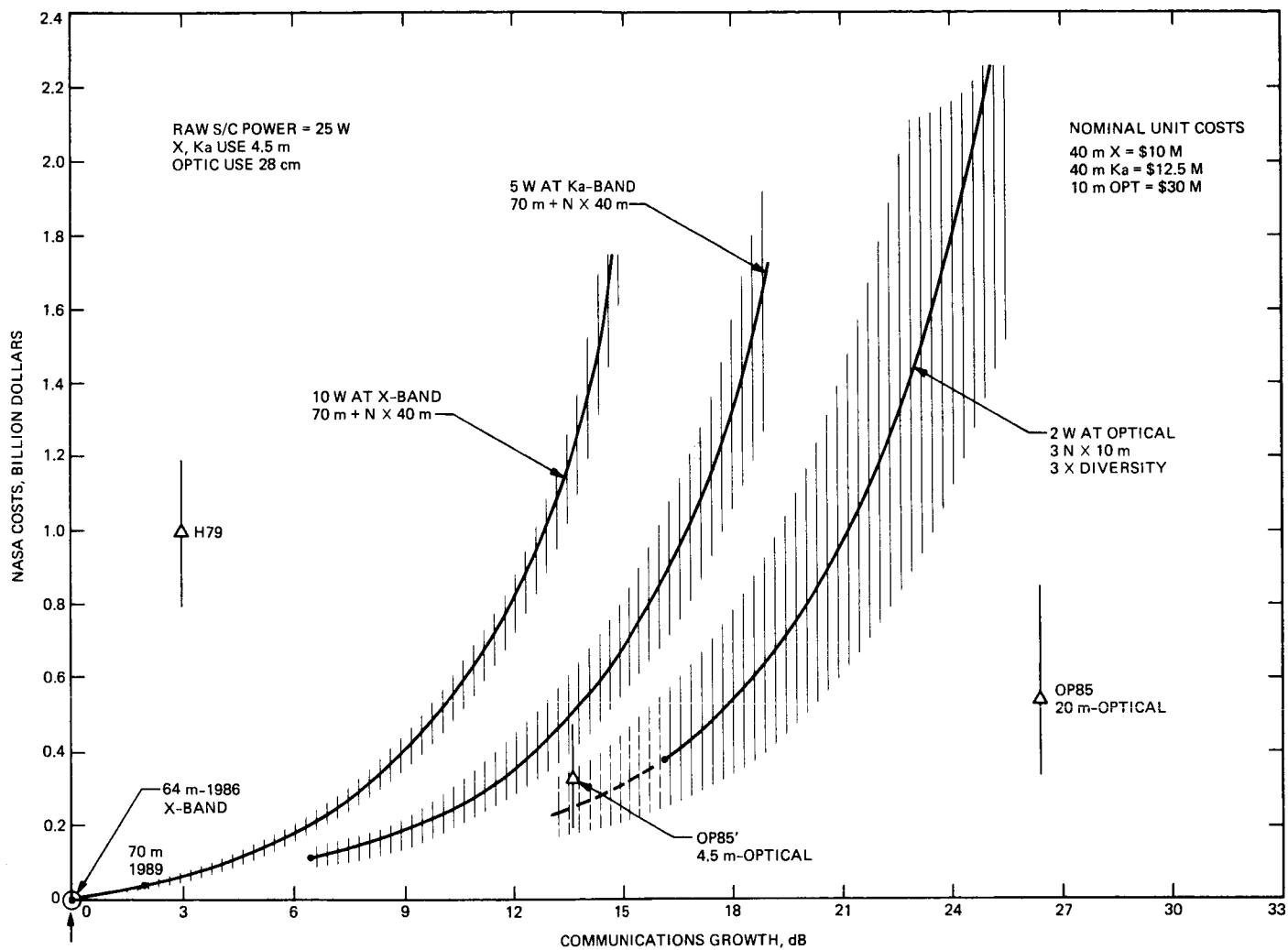
Fig. 1. Cost-performance comparison curves

Ka-Band (32-GHz) Performance of 70-Meter Antennas in the Deep Space Network

W. A. Imbriale, A. M. Bhanji, and S. Blank
Radio Frequency and Microwave Subsystems Section

V. B. Lobb, R. Levy, and S. A. Rocci
Ground Antenna and Facilities Engineering Section

Two models are provided of DSN 70-m antenna performance at Ka-band (32 GHz) and, for comparison purposes, one at X-band (8.4 GHz). The "baseline" 70-m model represents expected X-band and Ka-band performance at the end of the currently ongoing 64-m-to-70-m mechanical upgrade. The "improved" 70-m model represents two sets of Ka-band performance estimates (the X-band performance will not change) based on two separately developed improvement schemes: The first scheme, a "mechanical" approach, reduces tolerances of the panels and their settings, the reflector structure and subreflector, and the pointing and tracking system. The second, an "electronic/mechanical" approach, uses an array feed scheme to compensate for lack of antenna stiffness, and improves panel settings using microwave holographic measuring techniques. Results are preliminary, due to remaining technical and cost uncertainties. However, there do not appear to be any serious difficulties in upgrading the "baseline" DSN 70-m antenna network to operate efficiently in an "improved" configuration at 32 GHz (Ka-band). This upgrade can be achieved either by a conventional mechanical upgrade or by a mechanical/ electronic combination. An electronically compensated array feed system is technically feasible, although it needs to be modeled and demonstrated. Similarly, the mechanical upgrade requires the development and demonstration of panel actuators, sensors, and an optical surveying system.

I. Introduction

This report provides two models for 70-meter DSN antennas, each for Ka-band (32 GHz) performance, and, for comparison, one for X-band (8.4-GHz) performance. The first model, the "baseline" or 1990 version, models the expected antenna efficiency at the completion of the now ongoing 64-m-to-70-m upgrade. Though the upgrade for all three DSN 64-meter antennas is scheduled for completion prior to the 1989 Voyager encounter with Neptune, one upgrade may be delayed until after encounter — hence the label "1990 model." The model shows that the baseline 70-meter efficiency will exceed

75% at X-band, but be less than 30% at Ka-band at the 45-degree rigging angle (and substantially less as the antenna elevation differs from that angle).

The second model, the "improved" or 1995 model, describes the expected antenna efficiency after a further upgrade is made (around 1995) to accommodate Ka-band with improved efficiency. The 1995 model does not improve X-band performance noticeably, but increases the Ka-band performance in two ways — the maximum efficiency is increased to above 50%, and this efficiency is achieved over a wide range of elevation angles (from 10 degrees to 90 degrees).

The 1995 Ka-band model is based on two possible approaches to improving the antenna: (1) a "mechanical" approach that provides improved performance by stiffening the antenna and by improving panel settings using advanced surveying techniques, and (2) an "electronic/mechanical" approach that uses an array feed scheme to compensate for lack of antenna stiffness, and improves panel settings using microwave holographic measuring techniques.

The differences in the results of the two approaches are small in performance, though perhaps large in costs. However, cost data are preliminary. Therefore, although a 1995 model is described in this report, the actual approach to obtain that model is not specified.

This report reflects the best engineering judgment available during the summer of 1986. Subsequent analyses will reflect later judgment as Ka-band performance estimates continue to evolve.

II. Baseline, or 1990, Model

At the completion of the currently ongoing 64-m-to-70-m upgrade, the "baseline" antenna efficiency at both Ka-band and X-band is expected to be as shown in Table 1. Figure 1 shows gain versus elevation performance at the above two frequencies for the 70-m antenna.

The results in Table 1 and Fig. 1 are based on the following assumptions: (1) the surface panels will be rigged at a 45-degree elevation angle, (2) a 20-mph (32-km/h) wind is assumed to be blowing, and (3) a temperature range of 40°F (22°C) on the antenna structure is assumed. The surface panel rms manufacturing error is expected to be 0.008 in. (0.020 cm), and the accuracy of the panel settings is assumed to be within 0.015 in. (0.038 cm). Aperture phase error caused by small-scale atmospheric turbulence is included, but no atmospheric attenuation contribution is included (this is considered a space loss parameter). The assumed CONSCAN "tracking" accuracy of 0.002 degrees may represent a somewhat optimis-

tic view, making performance of the "baseline" antenna at 32 GHz as shown in Table 1 and Fig. 1 somewhat optimistic.

A. RF Loss (Degradation) Factors

The efficiency of a paraboloidal antenna is determined by both RF and mechanical factors, and by other factors (as shown in Table 1). Among the RF factors are those that cause loss when feed pattern characteristics are not ideal. These include the property of the energy source (feed) to illuminate only the reflectors while minimizing the energy that radiates elsewhere, and the property of the combined feed and subreflector to illuminate the parabola uniformly, making maximum use of the entire reflector surface. Ludwig (Ref. 1) and Potter (Ref. 2) have described these loss factors in detail.

B. Mechanical and Other Loss Factors

The uncertainty in the performance of the antenna at Ka-band due to the RF factors described above is outweighed by mechanical factors such as blockage and surface tolerance, and other factors such as pointing errors, wind, thermal effects, and atmospheric turbulence. The ground antenna effective aperture would remain essentially constant with increasing frequency were it not for aperture phase distortion — the so-called "Ruze" effect (Ref. 3) — which becomes serious as the standard deviation of these phase distortions approaches 1/12 of the wavelength. Mechanical deflections, manufacturing tolerances, wind, thermal effects, and small-scale atmospheric turbulence (Ref. 4) result in additional more-or-less random phase errors whose associated aperture efficiency loss factor, η_s, is given by

$$\eta_s = e^{-(4\pi\sigma/\lambda)^2}$$

where

σ = standard deviation of the one-half pathlength error

λ = wavelength

Using the above Ruze formula, various statistically independent factors responsible for phase error (and thus loss of aperture efficiency) can be calculated and tabulated. That is, for each independent factor i, there is an associated standard error, σ_i, and efficiency loss factor, η_{si}. Further, the total efficiency loss factor, η_{st}, is the product of the individual loss factors, and the total standard error, σ_t, is the rms of the individual standard errors. This is the basis for the development of Table 1.

The determination of Fig. 1 follows a similar line of reasoning, with the phase error being introduced by gravitational effects and atmospheric turbulence — a function of the eleva-

tion angle. This function is extracted from the works of Levy[1] and Potter (Ref. 4).

These figures differ slightly from a similar model developed last year (Ref. 5), essentially because of a higher expected efficiency, as less power is now expected to be contained in modes with azimuthal (m) variation unequal to 1 (i.e., modes that do not radiate on the antenna axis).

III. Improved, or 1995, "Mechanical" Model

The kernel of the "mechanical" approach for the improved model is to reduce tolerances of the reflector surface panels and panel settings, the reflector structure and subreflector, the pointing and tracking system, and those due to thermal effects. The net effect of these improvements is shown in Table 2 and Fig. 2.

A. Panels and Settings

Bonded panels (rather than the originally planned riveted panels with 0.008-in. rms tolerance) have now been fabricated to provide 0.004-in. (0.010-cm) rms tolerance. A precise setting of the reflector panels can further reduce the rms tolerance from the "baseline" 0.015 in.[2] to either 0.008 in. or 0.005 in. (0.0127 cm). However, the improved setting accuracy depends upon the development of an optical instrument that can be used at a 45-degree antenna elevation position to set the panels. The instrument must provide fully automatic readout, positioning, and ranging. It will require an operator to sight it on a corner cube optical target at the corner of the intersection of the panels. The expected accuracy of this instrument, including operator sighting resolution, will keep the error to within 0.008 in. at a range of 35 m. But an actual optical instrument that will achieve the above is not yet developed, although a prototype of this instrument is in use on the 70-m X-Band Upgrade Project (it must be used only in the zenith position). Corner cube optical targets and ranging instrument prototypes are currently being used, but the accuracy of the ranging and target holders needs to be improved before the accuracy stated above can be ensured.

Further precision setting of the reflector panels will require improvements to the Reflector Panel Controller so that individual reflector panel positions can be maintained within a pre-determined location to an accuracy of 0.005 in. (as a func-

tion of the antenna's elevation position). The key element of this improvement is a motorized actuator with a transducer that would position the four corners of its supporting panels. The planned approach is to use an open-loop system where either analytical or field-measured data can be used to position the panels, adjusting for effects of gravity on the antenna structure. A central computer would be used to control the 1800 actuators. These same actuators could also be used to compensate for thermal and wind distortions. Currently, the most difficult aspect of wind and thermal distortion compensation is the need for development of effective sensors or algorithms that derive the correct compensation movement for the actuators. Figure 3 is a drawing of a proposed actuator (in place on the antenna). Tailored after commercially available units, this device will fit within the limited space between the existing reflector structure and surface panels.

The gain improvement for the Reflector Panel Controller varies with antenna position. The improvement in gain is near zero at the rigging angle (where the panels are set by the Controller with the antenna at a 45-degree elevation position), and reaches a maximum improvement of 2.8 dB near the antenna zenith or horizon position.

B. Structure and Subreflector

A reinforced reflector structure would improve pointing and reduce wind distortion from 0.011 in. (0.028 cm) rms to 0.006 in. (0.015 cm) rms. In addition, the stiffening would reduce the pointing error caused by 20-mph winds to ±0.002 degrees.

Reduced tolerances are also possible from an improved subreflector. The improved subreflector would be two-pieced machine-cast aluminum, and symmetrically shaped. The improvement in the performance to the desired shape would be 0.006 in. rms versus the existing 0.010 in. (0.0254 cm) rms. The subreflector would be centered on the antenna boresite axis and shaped for optimum 32-GHz performance. The improved surface accuracy is readily achieved because the shape is a curve of revolution, and hence is easier to machine and verify. The current 70-m X-band subreflector is asymmetric in shape and requires (at least) a three-axis milling machine. In addition to the problem of the current subreflector being asymmetric, the vertical travel of the three-axis machine on which it is machined is not sufficient to cover the subreflector height. This requires resetting of the subreflector in order to complete the machining.

C. Pointing and Tracking

Antenna pointing and tracking improvements can be made in a series of progressive modifications, beginning with the tracking system, where there is a positive feedback correction

---

[1]Levy, R., JPL IOM 3556-84-092, Jet Propulsion Laboratory, Pasadena, CA, Oct. 10, 1984 (JPL internal memorandum); also F. D. McLaughlin, private communication.

[2]The baseline setting of 0.015 in. relies on holographic measurements taken at a 45-degree elevation angle.

system (CONSCAN). To be able to track and re-acquire a target at 32 GHz requires the following improvements for each stated accuracy (each improvement builds on the previous change):

(1) New 24-bit encoders and associated cabling and software to enable a 0.004-degree accuracy.

(2) The addition of a limited sensor package for measurements of antenna structure (for correction of temperature and gravity effects) to enable a 0.002-degree accuracy. Also needed is four-quadrant wind monitoring and a wind model for pointing correction.

(3) A sensor package to more completely monitor antenna gravity and temperature effects to enable a 0.001-degree accuracy. In addition, gravity panel corrections are required using the actuators for the panels and integrating a predictive wind correction model. Also, an adaptive control algorithm for enhanced antenna control is needed.

(4) An inertial reference to replace the present master equatorial system to provide 0.002-degree pointing acquisition. New systematic error correction software is also required.

D. Thermal

Providing thermal protection to the reflector structure and alidade can reduce the rms tolerance due to this effect from 0.010 in. to 0.005 in. The reflector structure would have insulated panels, enclosing the backside of the reflector structure. The resultant area would be separated into several plenum chambers, ventilated or temperature-controlled to provide uniform structure temperature changes. The alidade legs would be thermally insulated to provide for uniform temperature changes to the alidade structure.

IV. Improved, or 1995, "Electronic/Mechanical" Model

Table 3 and Fig. 4 give expected improved performance using "electronic/mechanical" upgrades, i.e., an array feed concept and advanced microwave holographic techniques. Cost factors are included. Figure 5 shows the array concept. The operational principle is described below.

A. Array Feed Concept

In principle, many aperture phasing errors may be compensated for electronically using an array feed. While the surface profile of the 70-m antenna may have sufficient accuracy at X-band, its surface distortion becomes a significant fraction of a wavelength at Ka-band, and the performance will sharply deteriorate. If these distortions can be compensated for by replacing the conventional single feed with an array feed, major performance improvements can be achieved at a fraction of the cost of building a new large-reflector antenna.

Rudge and Davis (Ref. 6) discussed the array feed concept for distorton correction and proposed the use of the Butler matrix and a means of providing adaptive excitation of the feed array. Their analytic and quantitative results, however, are limited to cylindrical reflectors having one-dimensional distortion profiles requiring only linear array feeds. Amitay and Zucker (Ref. 7) analyzed the use of planar array feeds for aberration correction in spherical reflectors. However, their synthesis procedure relies heavily on the circular symmetry of the feed-plane field distribution that exists in this case.

An algorithmic procedure has been developed at JPL to simultaneously provide electronic correction of systematic reflector distribution as well as electronic beam stabilization (Ref. 8). Although the computer code used in the study could be used to optimize general performance criteria (gain, sidelobe levels, and beam shape), this study focused only on gain maximization. For a given feed configuration, reflector f/d ratio, and distortion profile, the algorithm finds the optimum values of the individual feed excitations to maximize the reflector antenna gain in the desired direction.[3]

Once the array feed configuration has been determined, the array excitations must be adjusted to optimize performance. Adjustments to amplitude and phase can be accomplished at RF-with-digital phase shifters and variable power dividers or at some IF frequency using a device such as the Baseband Assembly (BBA). Since it is desirable to have as low a loss as possible, the front-end electronics (feed, low-noise masers) will be housed in a cryogenic front end. Significant trade-off studies have not yet been made to determine whether RF or IF beam forming would be preferred, but as a strawman design, the technique of using a BBA has been selected, and a block diagram is given in Fig. 5. It should be noted that since only seven elements are used, vernier beamsteering cannot be provided for a full ±1 beamwidth pointing error. Further studies are required to provide accuracte estimates of performance.

B. Microwave Holographic Measurements

Microwave holographic measurement techniques are expected to provide 0.015-in. settings by 1990, and 0.008-in.

---

[3]Numerical results were obtained for the "equivalent" single parabola case showing that good on-axis gain restoration can be achieved with as few as seven elements. For beam stabilization to ±1 beamwidth (BW), 19 elements are required. The dual-reflector case is presently under study.

settings by the 1995 upgrade. Further exploration is required in this area.

V. Conclusions

There do not appear to be any serious difficulties in upgrading the "baseline" DSN 70-m antenna network to operate efficiently in an "improved" configuration at 32 GHz (Ka-band). This upgrade can be achieved either by a conventional mechanical upgrade or by a mechanical/electronic combination. An electronically compensated array feed system is technically feasible, although it needs to be modeled and demonstrated. Similarly, the mechanical upgrade requires the development and demonstration of panel actuators, sensors, and an optical surveying system.

References

1. Ludwig, A. C., "Antenna Feed Efficiency," in *Supporting Research and Advanced Development*, Space Programs Summary 37-26, Vol. IV, pp. 200–208, Jet Propulsion Laboratory, Pasadena, CA, Apr. 30, 1964.

2. Potter, P. D., *The Aperture and Efficiency of Large Paraboloidal Antennas as a Function of Their Feed-System Radiation Characteristics*, Technical Report 32-149, Jet Propulsion Laboratory, Pasadena, CA, Sept. 25, 1961.

3. Ruze, J., "Antenna Tolerance Theory — A Review," *Proceedings of the IEEE*, Vol. 54, No. 4, pp. 633–650, Apr. 1966.

4. Potter, P. D., "64-Meter Antenna Operation at Ka-band," *TDA Progress Report 42-57*, pp. 65–70, Jet Propulsion Laboratory, Pasadena, CA, June 15, 1980.

5. Dickinson, R. M., *A Comparison of 8.415-, 32.0-, and 565646-GHz Deep Space Telemetry Links*, JPL Publication 85-71, Jet Propulsion Laboratory, Pasadena, CA, Oct. 15, 1985.

6. Rudge, A. W., and Davis, D. E. N., "Electronically Controllable Primary Feed for Profile Error Compensation of Large Parabolic Reflectors," *Proceedings of the IEEE*, Vol. 117, No. 2, pp. 351–358, Feb. 1970.

7. Amitay, N., and Zucker, H., "Compensation of Spherical Reflector Aberrations by Planar Array Feeds," *IEEE Transactions*, Vol. AP-20, No. 1, pp. 49–56, Jan. 1972.

8. Blank, S. J., and Imbriale, W. A., "Array Feed Synthesis for Correction of Reflector Distortion and Vernier Beamsteering," *TDA Progress Report 42-86*, pp. 43–55, Jet Propulsion Laboratory, Pasadena, CA, Aug. 15, 1986.

Table 1. 70-m antenna "baseline" performance estimates[a]

Contributor	X-Band (8.450 GHz)	Ka-Band (Baseline) (32 GHz)
Waveguide Loss	0.984	0.980
Dichroic Loss	0.992	–
Forward Spill	0.964	0.970
Rear Spill	0.997	0.997
Illumination	0.982	0.980
X-Polarization	1.000	1.000
Phase	0.989	0.980
Central Blockage	0.988	0.990
M ≠ 1 Modes	0.996	0.996
VSWR	0.991	0.990
Mesh Loss	0.999	0.998
RF Subtotal	0.887	0.887
Quadripod Blockage	0.9196	0.9196
Panels (0.008 in.)[b]	0.9948	0.9285
Setting (0.015 in.)[b]	0.9820	0.7704
Subreflector (0.010 in.)[b]	0.9919	0.8905
Gravity (30°)[c]	0.9948	0.9285
Thermal[b]	0.9920	0.8905
Wind (20 mph)[b]	0.9903	0.8691
Atmospheric Turbulence (30°) (Ref. 4)	0.9916	0.8864
Pointing (0.002°)[d]	0.9894	0.8568
Mechanical and Other Subtotal	0.854	0.320
Combined Efficiency Relative to 100%	0.758 −1.20 dB ± 0.2	0.284 −5.47 dB ± 0.5

[a]No atmospheric attenuation contribution is accounted for, as it is considered a space loss parameter.

[b]"64-Meter Antenna Rehabilitation and Performance Upgrade Project, Antenna Mechanical System," Ground Antenna and Facilities Engineering Section's E-Level Design Review (JPL internal document); also F. D. McLaughlin, private communication.

[c]Levy, R., JPL IOM 3556-84-092, Jet Propulsion Laboratory, Pasadena, CA, Oct. 10, 1984 (JPL internal memorandum); also F. D. McLaughlin, private communication.

[d]*Deep Space Network/Flight Project Interface Design Book*, Vol. I, TRK 10, Rev. B, p. 5, JPL Document 810-5, Jet Propulsion Laboratory, Pasadena, CA (JPL internal document), states "blind pointing or open loop pointing of 0.005 degrees and engineering estimates are made here for CONSCAN tracking to be 0.002 degrees (optimistic)."

Table 2. 70-m antenna "mechanical" upgrade performance estimates

Element	Improvement[a]	Quality
Reflector Panels	~0 dB	<0.004 rms
Subreflector	~0.3 dB	<0.006 rms
Quadripod	~0.1 dB	<2.5% blockage
Reflector Panel Setting	~0.8 dB	<0.008 rms
Reflector Panel Control	2.8 dB[b]	<0.005 rms
Thermal Control	0.4 dB	<0.005 rms
Reflector Stiffening (at ≤20-mph wind)	1.0 dB[b]	±0.002°
Tracking/Pointing Control	4.5 dB[b]	<0.001°/±0.002°

[a]Improvement over 70-m X-band.

[b]Varies with antenna position.

Table 3. 70-m antenna "electronic/mechanical" upgrade performance estimates

Contributor	Improvement Efficiency (a)	Improvement Efficiency (b)	dB	NRE	Per Antenna	Comments
RF	0.887	0.887	0.0	0	0	No more improvement after currently scheduled 64-m - 70-m upgrade
Quadripod Blockage	0.9196	0.9196	0.0	0	0	No more improvement after currently scheduled 64-m - 70-m upgrade
Mfg. Reflector Panels (0.008 in.)	0.9285	0.9285	0.0	0	0	No more improvement after currently scheduled 64-m - 70-m upgrade
Mfg. Subreflector (0.006 in. from 0.010 in.)	0.9591	0.8905	0.32	150K	800K	Cost estimate per Ground Antenna and Facilities Engineering Section for new subreflector
Setting (0.008 in. from 0.015 in.)	0.9285	0.7704	0.81		60K	Improved setting with holographic technique
Other Factors	0.8913	0.5458	2.13	2000K	3000K	Electronic phase error compensation using arrayed feed system for distortion correction
Combined Efficiency	0.600	0.284	3.26	$2150K	$3860K	
Relative to 100%	$-2.21 \text{ dB } ^{+0.0}_{-1.5}$	$-5.46 \text{ dB } \pm 0.5$		+2000K −500K	±1000K	

[a] After proposed upgrade (mechanical and electronic).

[b] At completion of currently scheduled 64-meter-to-70-meter upgrade.
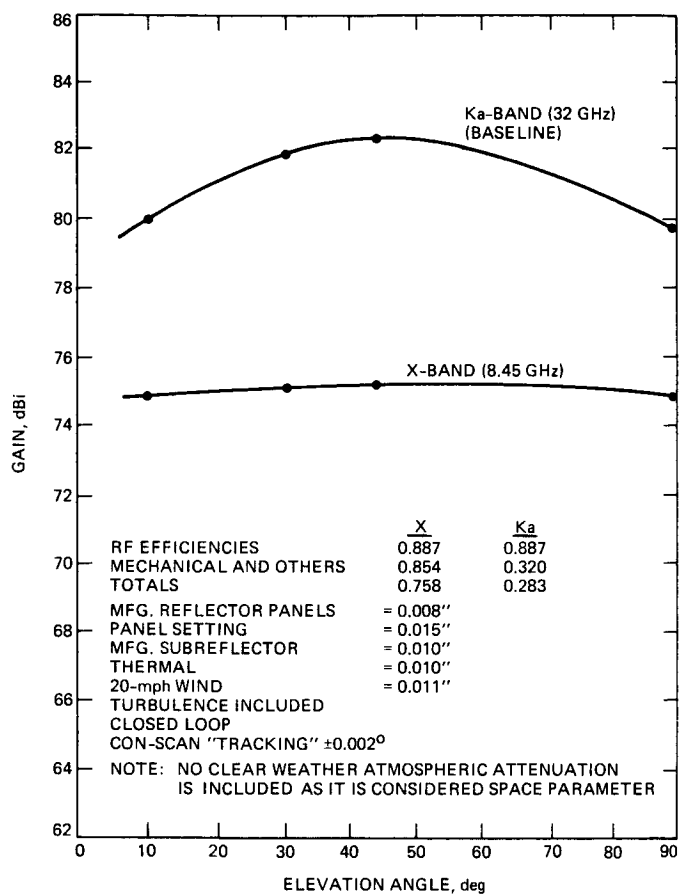
Fig. 1. Predicted antenna "baseline" performance at 64-m-to-70-m upgrade
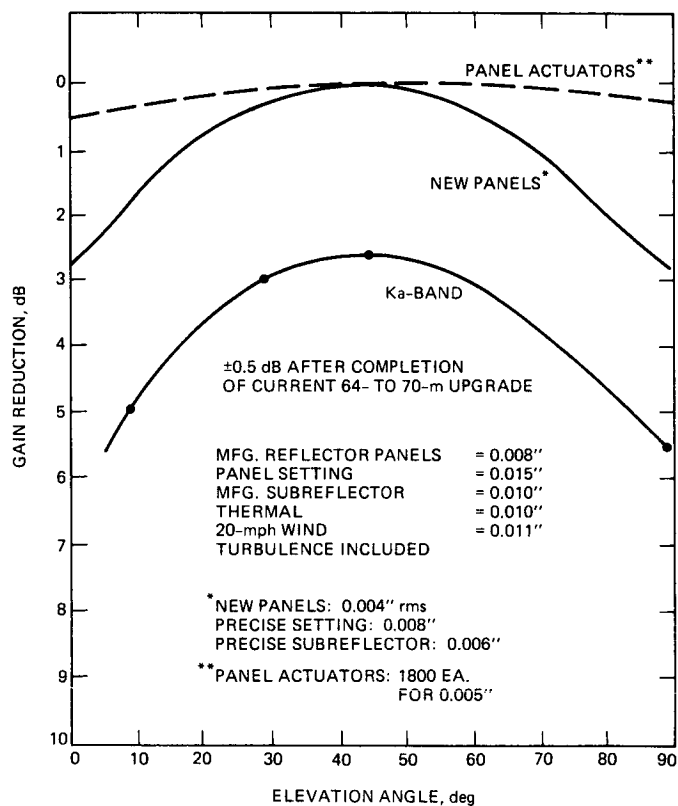
Fig. 2. Antenna gain loss caused by gravity effects for different elevation angles
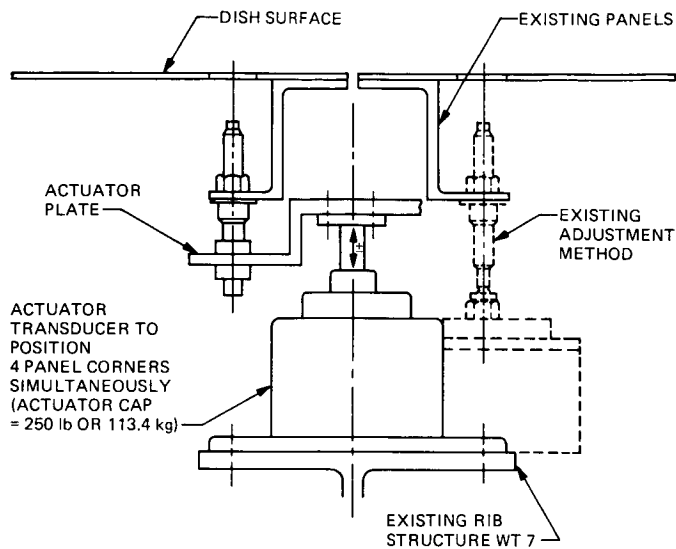
Fig. 3. Reflector panel adjustment, motorized actuator

PROPOSED 32-GHz 70-m ANTENNA EFFICIENCY IMPROVEMENT
MECHANICAL: 1. NEW SUBREFLECTOR 0.006"
            2.0 PRECISE SETTING 0.008"
ELECTRONIC: PHASE ERROR COMPENSATION
            USING ARRAY FEED SYSTEM
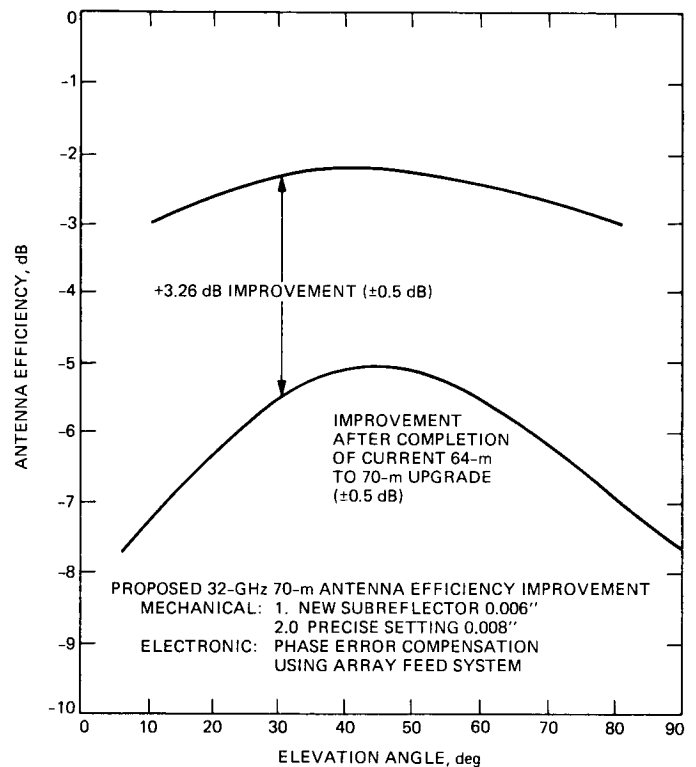
Fig. 4. 70-m antenna performance estimates, "electronic/mechanical" upgrade
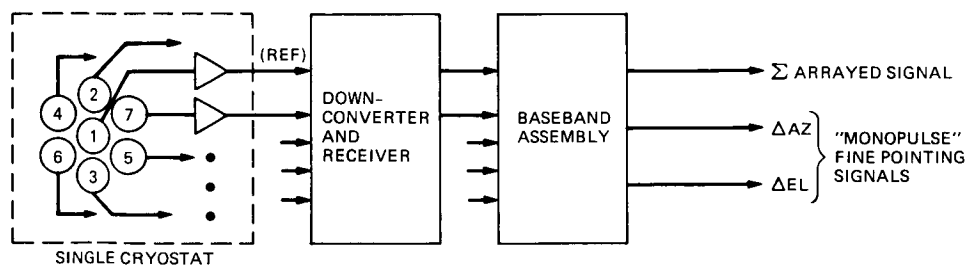
Fig. 5. Array feed concept

Models of Weather Effects on Noise Temperature and Attenuation for Ka- and X-Band Telemetry Performance Analysis

S. D. Slobin

Radio Frequency and Microwave Subsystems Section

Models that show the effects of weather on noise temperature and attenuation of deep space telemetry signals received by the Deep Space Network (DSN) at Ka- and X-band (32 GHz and 8.5 GHz) are developed. These models have been used to compare the performance of telemetry links at these two frequencies. The models build on an earlier 1982 model that used three months of water vapor radiometer measurements (31.4 GHz) at Goldstone, augmented with one year of radiosonde measurements made at Edwards Air Force Base. This 1986 model accounts for annual variations of rainfall and extends to a model for Canberra, Australia, and Madrid, Spain. The results show, for example, that at Ka-band, 30-degree elevation angle, Goldstone weather adds less than 23 ± 2 K to the system temperature 80% of the time, while Canberra or Madrid weather adds less than 32 ± 5 K 80% of the time. At X-band, the comparable numbers are 5.1 ± 0.2 K and 5.7 ± 0.4 K. A simple analysis shows a substantial telemetry system signal-to-noise ratio advantage when operating at Ka-band compared to X-band.

I. Introduction

X- and Ka-band (8.5 GHz and 32 GHz) weather-effects models are developed for use in a comparison of DSN telemetry link performance at these two frequencies (Ref. 1). A large number of independent weather models and statistical studies exist at various frequencies, but they are unrelated, of different data types, and the data are taken under varying and different experimental conditions. A direct comparison of any two such models would not allow satisfactory comparison of link performances at the two frequencies of interest. Therefore, new models have been created.

A Ka-band noise temperature and attenuation model developed at JPL was used as a starting point (Ref. 2). This model is based on 31.4-GHz water vapor radiometer measurements carried out at Goldstone, augmented by one year's radiosonde measurements made at Edwards Air Force Base. That Ka-band model for Goldstone was developed in such a way as to ensure a worst-case analysis, i.e., the model may be considered as typical of a "worst-year." If this is considered to be a 2-sigma case, then only one year in fifty would have more water vapor, clouds, and rain, and thus more atmospheric attenuation and higher noise temperatures.

With this limited model (the radiometer measurements were only made during a three-month period, the winter), the problem then existed as to how to model average and best years at Goldstone, and then how to model the overseas DSN stations at Madrid and Canberra.

II. Methodology of Weather Modeling

The derivative models (from Goldstone, worst-year) were developed on the basis that when one year has less total rain than another, it is because it rains fewer total minutes rather than the same number of minutes at a lower rate. It is recognized that the actual situation is a combination of these two conditions. Thus, in modeling for average and best cases, using the worst-case as a starting point, adjustments were made to the cumulative distribution (CD) value for a given value of noise temperature. For example, if an average year had half the rain of a worst-case year, then the noise temperature value at the 98% CD value (for worst-case) was used for the noise temperature value at CD = 99%. Or, in other words, a noise temperature value (corresponding to a particular rain rate) which is exceeded 2% of the time in a worst-case year is exceeded only 1% of the time in an average year.

Table 1 shows the reported Goldstone Ka-band, worst-month weather effects model as presented in Ref. 2. This table may be interpreted to read, for example: "95% of the time the total Ka-band atmospheric noise temperature is 35 kelvins or less," subject to the conditions stated in the table description. Figure 1 shows the graphical display of the cumulative distribution. As the distribution of values above 99% is important to illustrate the occurrence of rare events, "probability paper" is used in these presentations.

The first step in modeling the Goldstone average and best cases was to determine the relationship among the worst, average, and best years on the basis of some measure or factor. A study of Los Angeles rainfall records for the 106-year period 1877–1983 showed that compared to an average year, the rainiest years had about twice the average rainfall, and the driest years had about one-third the average rainfall. For Los Angeles, the three rainfall values are, approximately, 5, 15, and 30 inches (12.7, 38.1, and 76.2 cm) per year. The distribution is not Gaussian. Forty-four years had above-average rainfall, and sixty-two years had below average. The lowest rainfall year had 4.85 inches (12.32 cm), the average had 15.11 inches (38.38 cm), and the highest had 38.18 inches (96.98 cm). Values greater than +/– "two-sigma" were discarded in this qualitative analysis.

Two other qualitative arguments were brought to bear in this model development. First, that on a worldwide average, there is 50% cloud cover (by area or time); and, secondly, that during the 4400 non-cloudy, non-rainy hours of the year, the clear-sky attenuation and noise temperatures have the same distribution, independent of wet or dry years. The latter argument results in an *a priori* decision that the cumulative distributions would be identical up to the 50% level, and that only above that level would they diverge into best, average, and worst cases.

The arguments used to develop best and average cases (for Goldstone) from the worst-case model can best be described as heuristic. To link the rainfall factors (1/2 and 1/6 compared to maximum) with the requirement at 50% CD, a sliding scale (based on CD) was developed which was applied to the cumulative distribution value for the particular noise temperature value given in Table 1. Thus, for example, to derive a CD value for average-year at Goldstone for 69 kelvins, the CD value of 99.5% is changed to 99.75%. That is, in a worst-year where the noise temperature value of 69 kelvins is exceeded 0.5% of the time, in an average-year it is exceeded only 0.25% of the time. Similarly, in a best-year, it is exceeded only 0.083% of the time (1/6 of 0.5%) and the CD becomes 99.917% at 69 kelvins. Note that the ratios of the exceedance percentages are the same as the rainfall ratios discussed earlier. The effect of the changing exceedance values is to move the cumulative distribution curves "up" toward 100% when moving from worst to best cases. The sliding CD scale is used to move noise temperature points half-way or 5/6 way (for average and best, respectively) at CD's near 100% (the CD region in which rain is an important contributor). Points with lower CD values are moved less (in the cloud region) until at 50% CD, the worst, average, and best curves intersect. Below 50% only one CD curve exists, the one shown in Fig. 1.

Now that the three distribution curves for Goldstone can be generated, the question remains as to the modeling of Canberra and Madrid Ka-band statistics. Again, rainfall is used as the factor by which one site or condition may be compared to another. The Goldstone yearly rainfall average is 3.5 inches (8.9 cm). Canberra and Madrid average 23.0 and 19.6 inches (58.4 and 49.8 cm), respectively. For the purposes of this report, it was decided to treat Canberra and Madrid as similar in terms of weather. Year-to-year variations in rainfall at either station are certainly greater than the difference between the stations, and they are certainly more like each other in vegetation and appearance than when compared to Goldstone (3.5 inches or 8.9 cm) or New York (40 inches or 102 cm). The average rain for the two overseas sites is then 21.3 inches (54.1 cm), a factor of 6.08 higher than the Goldstone rain total.

To generate the Ka-band average-year curve for Canberra/ Madrid, the exceedance percentages for the Goldstone curve (1%, 5%, etc.) are simply multiplied by a factor of 6 to reflect the rainfall ratio. In this method, it is also assumed that the cloudiness factor (the product thickness and occurrence) is greater by the same factor. There is no sliding CD scale as in the calculation of cases at a particular location. As an example, a 98% CD for Goldstone becomes 88% for Canberra/Madrid, 95% becomes 70%, and so forth. Clearly this cannot go on for all values (83% does not go to 0%). It is known that the theoretical clear-dry, oxygen-only 0% CD value of noise tempera-

ture for Canberra/Madrid (average altitude = 0.730 km MSL) at 30-degree elevaton angle is 10.5 kelvins. Thus the average noise temperature curve must pass through this point at 0% CD. Judicious choice of plotting techniques (which invoked a natural requirement that the CD curves be smooth, orderly, and separated) resulted in a set of 6 curves for Goldstone and Canberra/Madrid which give the range of total atmospheric noise temperature at all DSN sites over what is believed to be an approximately 2-sigma range of yearly-average weather conditions. The curves thus generated are shown in Fig. 2, for a 30-degree elevation angle. It should be noted that the curves deviate slightly from the requirement of equality below 50% CD.

III. Models for Attenuation and Elevation Angle Effects

In order to generate noise temperatures for a range of elevation angles, it is first necessary to create an attenuation model. This attenuation model (at 30-degree elevation) can be moved by 1/sin (elevation) to yield attenuation models at other elevation angles. The noise temperatures themselves cannot be so modeled.

First, for 30-degree elevation, an attenuation model for CD's on the six noise temperature curves is generated from

$$ATTN30 = 10 * LOG ((TP/(TP-T30))$$

where

TP = physical temperature of the atmosphere, typically 280 kelvins

T30 = noise temperature values at 30-degree elevation for each CD

Table 2 gives the 30-degree elevation values for noise temperature and attenuation models at Goldstone and Canberra/Madrid. The attenuation values can be modeled by 1/sin(elev) to create attenuation values at other elevation angles. New noise temperature values can then be created from these attenuation values by

$$TEMP = TP * (1-1/L)$$

where

TP = 280 K

L = 10 ** (ATTN/10)

ATTN = attenuation value in dB at another elevation angle, modeled as described above

In this way complete noise temperature and attenuation values for all locations, conditions, elevation angles, and CDs can be created from the initial six Ka-band noise temperature curves at a 30-degree elevation angle as described above. The multitude of numbers presented in the Ka-band study are not presented here.

IV. Modeling Noise Temperature and Attenuation Values at X-Band

In order to make the comparison of telecommunication performance at X- and Ka-bands, another complete set of noise temperature and attenuation values must be created for X-band. It is acknowledged that the effects of water vapor, clouds, and rain for the two frequencies 8.5 and 32 GHz hold an approximately frequency-squared relationship with one another. This ratio is 14.2. Clearly, if extrapolation by frequency is to be done, it is much better to go down in frequency than up. This is because errors in the higher frequency model are reduced by a factor of 14 when creating a model at the lower frequency. In the other direction, small errors at the lower frequency are greatly magnified when moving up in frequency.

As only the "wet" components (water vapor, clouds, rain) have a frequency-squared relationship between X- and Ka-bands, the "constant" oxygen value at Ka-band must be removed first. The modeling will be first done with the attenuation values, so, for example, the 0% CD values (oxygen only) of 0.158 and 0.166 dB at 30-degree elevation are subtracted from all attenuation values (c.f. Table 2). Other clear-air oxygen-only values will apply at other elevation angles. The remaining wet-component attenuation values are divided by 14.2 to create the X-band wet-component attenuation values. The X-band clear-air oxygen-only values are then added back in to create the total X-band attenuation values. For the example presented, at 30-degree elevation, these clear-air values are 0.064 dB, approximately, for all three stations. Then, in a process as described above, X-band noise temperature values can be calculated from the attenuations to create a complete X-band noise temperature and attenuation model for all stations, conditions, and elevation angles. The 30 degree elevation results are given in Table 3.

V. Integration of the Atmospheric Effects Model With the Telecommunications Performance Analysis Model

As developed a number of years ago, DSN telecommunications performance analysis programs contain a couple of awkward but correct steps in the calculation of antenna gain and system noise temperature at X-band. Curves of X-band antenna

gain and efficiency created by observation of known radio sources contain within them the elevation angle effects of clear-sky atmospheric attenuation. Presumably, at the time of measurement, it was not possible to account for the atmospheric effects and they were retained in "antenna gain." It is incorrect and misleading (especially at higher frequencies) not to account separately for the atmospheric effects and just lump them in with antenna gain. If this is done, without adequate monitoring of temperature, pressure, and relative humidity during clear-sky conditions, it will not be possible to adequately calibrate antenna performance. Day-to-day atmospheric variation will mask any subtle changes in antenna performance obtained by panel adjustment, pointing improvement, subreflector positioning, etc. The ground noise contribution from quadripod scatter and rear spillover with decreasing elevation angle are included in a baseline curve of "clear weather receiving system noise temperature increase for non-zenith elevation angles.[1]" For X-band, the actual variable clear-sky noise temperature differences from those included in the gain and efficiency curves are probably not enough to create large errors in a model of total system noise temperature. If these curves were used at Ka-band, large noise temperature errors would probably result. Atmospheric effects, antenna gain, and ground noise contribution must be determined separately from one another; and during any antenna calibrations at frequencies higher than X-band, an adequate monitor of atmospheric parameters must be maintained at all times.

---

[1]"DSN Telecommunications Interfaces, Atmospheric and Environmental Effects," TCI-40, Dec. 1, 1983, in *Deep Space Network/Flight Project Interface Design Handbook, Volume I: Existing DSN Capabilities,* Revision D, JPL Internal Document 810-5, Jet Propulsion Laboratory, Pasadena, California.

Because of the awkward construction of the telecommunications analysis programs, the 0% CD values presented here in all models must be subtracted from all noise temperature and attenuation models so that the remaining atmospheric model is relative to the clear-sky baseline at the elevation angle of interest. The actual clear-sky values of attenuation and noise temperature are included in elevation angle curves of antenna gain and system noise temperature, respectively.

VI. Conclusions

The X- and Ka-band noise temperature models developed here allow a direct comparison of telemetry system performance at the two frequencies. All other things being equal (e.g., spacecraft transmitter power, antenna efficiency, antenna pointing), the gain advantage obtained by operating an antenna at Ka-band rather than X-band (proportional to frequency-squared, 11.5 dB) more than outweighs the additional atmospheric attenuation and noise temperature increase at the higher frequency. As an example, using the atmospheric models developed here and following the argument presented in Ref. 2, it can be shown that for Canberra/Madrid, 30-degree elevation angle, 80% of the time the signal-to-noise ratio (SNR) advantage of Ka-band over X-band will be 8 dB or more. This represents a substantial improvement in DSN telemetry system performance and warrants further detailed measurements and analyses of Ka-band atmospheric effects. Scattered, short-term measurements by JPL at Ka-band tend to support the models presented here; however, greater confidence in these models (or development of updated ones) will depend on long-term measurements (typically five to ten years) at all DSN antenna locations.

References

1. Koerner, M. A., "Relative Performance of 8.5-GHz and 32-GHz Telemetry Links on the Basis of Total Data Return per Pass," *TDA Progress Report 42-87*, pp. 65-80, Jet Propulsion Laboratory, Pasadena, California, November 15, 1986.

2. Clauss, R. C., Franco, M. M., and Slobin, S. D., "K_A-Band Weather-Dependent System Performance Estimates for Goldstone," *TDA Progress Report 42-71*, pp. 60-65, Jet Propulsion Laboratory, Pasadena, California, November 15, 1982.

Table 1. Ka-band noise temperature and attenuation model for Goldstone, worst-month, 30-degree elevation angle (Ref. 2)

Cumulative Distribution, %	Total Atmosphere Noise Temperature, K	Total Atmosphere Attenuation,[a] dB
99.5	69	1.07
99	53	0.82
98	43	0.67
95	35	0.54
90	29	0.45
80	25	0.39
50	19	0.29
0[b]	9.84[c]	0.15

[a] Derived from noise temperature by 0.1 dB = 6.45 K.
[b] Clear-dry, typical of a very cold winter night.
[c] Theoretical oxygen-only.

Table 2. Ka-band noise temperature and attenuation models for Goldstone and Canberra/Madrid for 30-degree elevation angle

CD, %	Goldstone Best	Goldstone Avg	Goldstone Worst	Canberra/Madrid Best	Canberra/Madrid Avg	Canberra/Madrid Worst
			Noise Temperature, K			
99.9	65.	120.	240.	260.	270.	279.
99.8	51.	80.	175.	180.	215.	260.
99.5	41.	53.	69.	75.	150.	220.
99.	36.	44.	53.	57.	120.	180.
98.	32.	37.	43.	46.	62.	120.
95.	27.	31.	35.	37.	46.	55.
90.	24.	26.	30.	33.	38.	45.
80.	21.	23.	25.	29.	32.	37.
70.	19.	20.5	22.	26.	28.5	32.
50.	17.	18.	19.	23.	25.	27.
20.	13.5	14.	14.5	16.5	17.5	18.5
0.	10.	10.	10.	10.5	10.5	10.5
			Attenuation, dB			
99.9	1.147	2.430	9.451	11.461	14.472	24.472
99.8	0.873	1.461	4.260	4.472	6.342	11.461
99.5	0.688	0.911	1.229	1.354	3.332	6.690
99.	0.598	0.742	0.911	0.989	2.430	4.472
98.	0.527	0.616	0.724	0.779	1.087	2.430
95.	0.440	0.510	0.580	0.616	0.779	0.950
90.	0.389	0.423	0.492	0.545	0.633	0.761
80.	0.339	0.372	0.406	0.475	0.527	0.616
70.	0.305	0.330	0.355	0.423	0.466	0.527
50.	0.272	0.289	0.305	0.372	0.406	0.440
20.	0.215	0.223	0.231	0.264	0.280	0.297
0.	0.158	0.158	0.158	0.166	0.166	0.166

Table 3. X-band noise temperature and attenuation models for Goldstone and Canberra/Madrid for 30-degree elevation angle

CD, %	Goldstone Best	Goldstone Avg	Goldstone Worst	Canberra/Madrid Best	Canberra/Madrid Avg	Canberra/Madrid Worst
			Noise Temperature, K			
99.9	8.5	14.1	38.9	50.4	61.3	94.1
99.8	7.3	9.9	21.9	22.7	30.4	50.4
99.5	6.5	7.5	8.9	9.4	17.9	31.8
99.	6.1	6.7	7.5	7.8	14.1	22.7
98.	5.7	6.1	6.6	6.8	8.2	14.1
95.	5.4	5.7	6.0	6.1	6.8	7.6
90.	5.1	5.3	5.6	5.8	6.2	6.8
80.	4.9	5.1	5.2	5.5	5.7	6.1
70.	4.8	4.9	5.0	5.2	5.4	5.7
50.	4.6	4.7	4.8	5.0	5.2	5.3
20.	4.4	4.4	4.4	4.5	4.6	4.7
0.	4.1	4.1	4.1	4.1	4.1	4.1
			Attenuation, dB			
99.9	0.134	0.224	0.649	0.861	1.073	1.779
99.8	0.114	0.156	0.353	0.368	0.500	0.861
99.5	0.101	0.117	0.140	0.148	0.287	0.524
99.	0.095	0.105	0.117	0.122	0.224	0.368
98.	0.090	0.096	0.104	0.107	0.129	0.224
95.	0.084	0.089	0.094	0.096	0.107	0.119
90.	0.080	0.083	0.088	0.091	0.097	0.106
80.	0.077	0.079	0.082	0.086	0.089	0.096
70.	0.074	0.076	0.078	0.082	0.085	0.089
50.	0.072	0.073	0.074	0.079	0.081	0.083
20.	0.068	0.069	0.069	0.071	0.072	0.073
0.	0.064	0.064	0.064	0.064	0.064	0.064
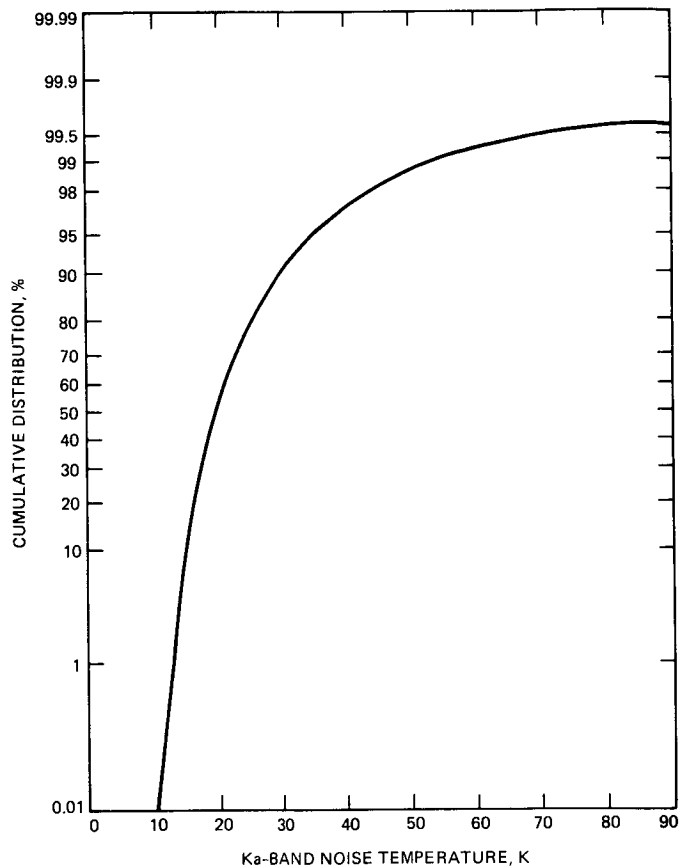
Fig. 1. Ka-band atmospheric noise temperature statistics: worst case, Goldstone, 30-deg elevation angle (Ref. 2)
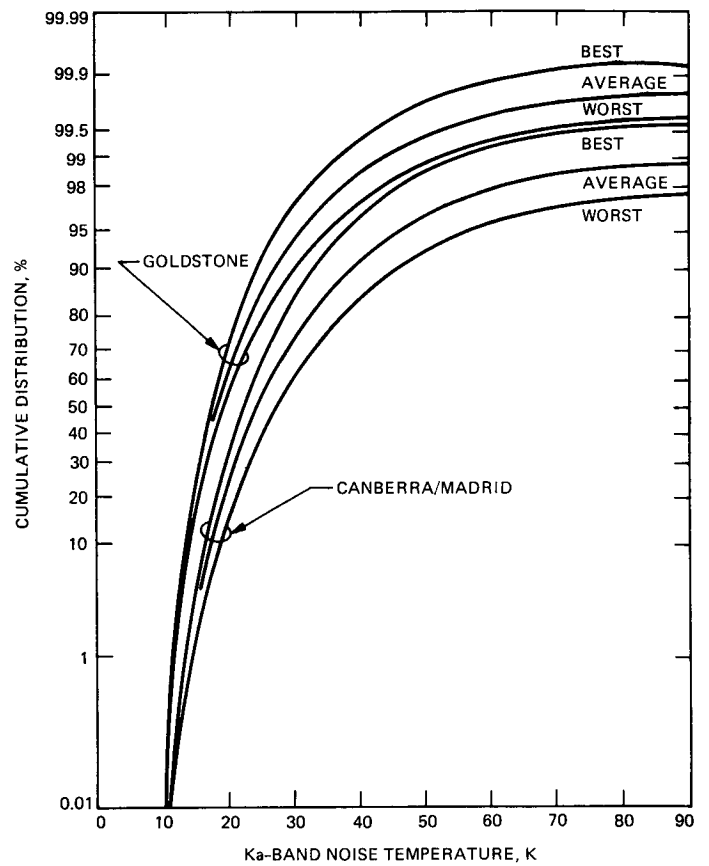
Fig. 2. Ka-band atmospheric noise temperature statistics: all sites, 30-deg elevation angle

A Ka-Band (32 GHz) Beacon Link Experiment (KABLE) With Mars Observer

A. L. Riley
Spacecraft Telecommunications Equipment Section

D. M. Hansen and A. Mileant
Telecommunications Systems Section

R. W. Hartop
Radio Frequency and Microwave Subsystems Section

A proposal for a Ka-Band (32 GHz) Link Experiment (KABLE) with the Mars Observer mission has been submitted to NASA. The experiment will rely on the fourth harmonic of the spacecraft X-band transmitter to generate a 33.6 GHz signal. The experiment will rely also on a DSN receiving station equipped to simultaneously receive X- and Ka-band signals. The experiment will accurately measure the spacecraft-to-Earth telecommunication link performance at Ka-band and at X-band (8.4 GHz).

I. Introduction

This article describes a proposal to implement the Mars Observer spacecraft (MO) with a Ka-band beacon to enable a Ka-Band Link Experiment (KABLE). The experiment will measure and document the Ka-band (32 GHz) deep space communications link advantage with respect to the X-band (8.4 GHz) link. Implementation of a Ka-band beacon in the Mars Observer spacecraft and a simultaneous X- and Ka-band receiving capability in a DSN R&D receiving station is required. Added benefits of the Ka-band beacon include the ability to calibrate plasma effects not possible with a single frequency radio link and the enhancement of gravitational wave experiments, relativistic bending experiments and spacecraft naviga-

tion. Radio science experiments are coordinated with the Mars Observer radio science team and are not a part of the proposal.

Propagation measurements and analyses show that a significant advantage can be achieved by using Ka-band instead of X-band for deep space communications and navigation. The performance advantage in the communications link can be as large as a factor of 10 and needs to be quantified with the precision required for future mission planning. Previous measurements of the effects that degrade deep space communications link performance suffered due to significant errors in the absolute accuracy of the separate evaluations. A simultaneous relative measurement of Ka-band and X-band

link performance will yield the necessary high accuracy results.

II. Experiment Objectives

The primary engineering objective of this experiment is to accurately measure and document the Ka-band spacecraft-to-Earth telecommunications link (downlink) performance with respect to the X-band downlink performance.

Parameters of interest include:

(1) Effects of ground antenna pointing and efficiency as a function of elevation angle, wind, irregular atmospheric refraction and thermal gradients experienced at sunrise, sunset and during a variety of weather conditions.

(2) Effects of water vapor and water droplets (clouds and rain) along the propagation path at various ground antenna elevations.

(3) Effects of solar corona and solar wind.

Documentation of these telecommunications link parameters will provide accurate inputs for future deep space mission planning. This will enable future flight projects to maximize the benefits of using improved telecommunications link capabilities at Ka-band.

It is expected that this experiment will confirm the prediction that during 90% of the time the Ka-band telemetry link performance will exceed the X-band telemetry link performance by a factor of 6 to 10. The 90% time period is based on current knowledge of the effects of weather including wind, clouds, and rain. Recent data obtained during the Voyager Uranus encounter have added to evidence that shows the DSN X-band weather model to be overly pessimistic. This experiment will significantly increase knowledge about the effects of weather and enable the improvement of the DSN weather models both for X-band and Ka-band.

The frequency and timing stability of the entire link (Earth based transmitter to spacecraft and back to the Earth based receiver) will be measured and is expected to be near 1 part in 10^{15}. Use of a Ka-band link offers a factor of 16 reduction in the sensitivity to plasma effects over X-band. Use of simultaneous X- and Ka-band offers the additional capability to calibrate the plasma and thereby reduce its effect even more. This can result in a factor of at least 10 improvement in stability and calibration accuracy over existing systems and may lead the way to new discoveries through an improved radio science capability.

III. Experiment Description

The proposed experiment requires the use of the fourth harmonic of the spacecraft X-band downlink signal thereby providing a Ka-band downlink to a DSN R&D Earth station. In order to minimize cost and spacecraft impact, several design approaches will be investigated. Selection of the specific approach to be implemented will be made by the Mars Observer Project office. Design approaches include:

(1) Traveling-Wave Tube Amplifier (TWTA) Generation of the fourth harmonic. This approach utilizes the fourth harmonic of the X-band downlink signal as generated by the TWTA; it is believed to be the least expensive and least intrusive to the spacecraft. Analysis is extremely difficult because the Ka-band signal must propagate through the X-band waveguide and components which are overmoded at Ka-band. Usefulness of this approach will be investigated for comparison with the other, better understood approaches in consideration of its possible use.

(2) Dual-Frequency Feed with Varactor Diode Quadrupler. This approach uses a dual-frequency feed at the primary focal point of the antenna with a varactor diode assembly to produce the fourth harmonic of the X-band downlink signal. The details of the multiplier scheme are covered in the third approach. The multiplier package is attached to the focal point feed assembly, adds no additional X-band waveguide or components, uses the same amount of the X-band downlink power, and may add less mass and cost less than the third approach. A better understanding of the spacecraft antenna feed system is needed in order to evaluate this approach accurately.

(3) Separate Cassegrain Ka-Band Feed with Varactor Diode Quadrupler. This approach is best understood and is being used here for scheduling and budgeting purposes; a detailed description follows.

The Ka-band beacon will consist of a passive frequency multiplier driven by an X-band signal coupled from the power amplifier output. The multiplier will produce a signal at the fourth harmonic (approximately 33.6 GHz) of the X-band downlink. This Ka-band signal will then be fed to the MO high-gain antenna for transmission to Earth. The Ka-band signal will be coherent with the X-band downlink signal and, when the X-band signal is phase modulated, will provide a modulated signal with four times the modulation index of the X-band signal. This will impact the resultant Ka-band carrier suppression and the effect on the experimental link performance. The Ka-band carrier signal-power-to-noise-spectral-density ratio (P_c/N_o) versus time for 20 mW transmitter power

is shown in Fig. 1 for a DSN 34 meter high-efficiency antenna during clear, dry weather. The time scale in these figures begins at Mars orbit insertion (August 1991). Typical link parameters are given in Table 1.

IV. Flight Hardware

The RCA Mars Observer spacecraft is shown in Fig. 2 with the high-gain antenna located on an extended boom. The antenna is mounted to the boom through a two-axis gimbal containing X-band waveguide rotary joints. The reflector is 1.0 meter in diameter, and the spacecraft transmitter produces 40 W of X-band power at the TWTA output. Figures 3 and 4 show a concept of the Ka-band experiment hardware mounted to the high-gain antenna. A block diagram is shown in Fig. 5.

The multiplier drive is obtained from a crossguide coupler which couples approximately 200 mW of X-band signal to the multiplier assembly, reducing the transmitted X-band signal by 1%. X-band to Ka-band conversion efficiency of 25% or more is theoretically possible but to keep the design simple and cost low a minimum conversion efficiency of 10% will be considered to produce a minimum output power of 20 mW at Ka-band. The quadrupler will consist of a cascade of two frequency doublers. Each doubler will consist of a varactor diode with an input low-pass matching network and output band-pass matching network. The varactors will be self-biased such that bias current is generated by rectification of the RF signal by the varactor. Thus, no bias supply voltage is required from the spacecraft.

The Ka-band signal is fed from the multiplier through the back of the high-gain antenna reflector to a Ka-band feed. The signal from the feed is reflected from a dichroic Cassegrain subreflector to illuminate the main reflector. The X-band signal is transmitted through the subreflector. The mass of the additional hardware is estimated to be about 1 kilogram.

Thus, the RCA proposed antenna design is modified to provide the following interfaces with the Ka-band experiment:

(1) An X-band 20 dB crossguide coupler to provide multiplier input signal.

(2) A mounting bracket on the back of the reflector for the frequency multiplier.

(3) A port through the back of reflector for Ka-band feed.

(4) A mounting bracket for the dichroic subreflector.

A number of issues require further investigation. These include an understanding of the mass handling margin of the high-gain antenna gimbal and boom system, and the impact of the high-gain antenna thermal environment upon the frequency multiplier operation.

The equipment will be designed to enable verification of proper multiplier operation through the use of a bias voltage test point located on the multiplier assembly. Measurement of this bias voltage will be done manually; this measurement is not required after launch.

V. Ground Support Equipment

A new 34-meter antenna for Research and Development at the Venus Site, Goldstone, California, is planned for completion in 1990. The DSN Advanced Systems Program supports a number of ongoing development efforts (RTOPs) that will equip this antenna to simultaneously receive 8.4 GHz (X-band) and 33.6 GHz (Ka-band) signals. This ground station will provide high-aperture efficiency and ultra-low-noise receiving systems with frequency stability of approximately one part in 10^{15}. The station will include a multifrequency beam waveguide feed system, X- and Ka-band ultra-low-noise amplifiers (cryogenically cooled masers and high-electron mobility transistors), hydrogen maser frequency standards, fiber optic distribution systems, an ultra-stable 100 kW X-band (7.145–7.190 GHz) transmitter, receivers, a spectrum analyzer with –180 dBm sensitivity and one hertz resolution, and other advanced signal, and data processing and recording equipment.

This experiment complements and enhances efforts within the DSN Advanced Systems Program to develop and demonstrate Ka-band station technology. The Ka-band link experiment will be closely coordinated with the DSN Advanced Systems Program so that no additional resources will be needed to develop, install, use, and maintain the ground support equipment.

VI. Mission Operations

The objectives of this experiment can be accomplished without causing any change to the Mars Observer Project mission plan. Tracking functions at the R&D Venus station will be coordinated with the Mars Observer mission manager. Scheduling of the Venus station will be accomplished and supported by the DSN Advanced Systems Program as has been and continues to be the approach used for other research activities in the DSN.

VII. Analysis and Report

The analysis of data acquired at the R&D Venus site in support of this experiment will quantify the following parameters:

(1) The Ka-band downlink performance with respect to the X-band downlink performance.

(2) Ground antenna pointing and efficiency as a function of elevation, wind, irregular atmospheric refraction and thermal gradients.

(3) Effects of water vapor and water droplets along the propagation path at various ground antenna elevations.

(4) Effects of solar corona and solar wind.

(5) Total system "turn-around" frequency stability.

VIII. Conclusions

The proposed experiment can provide an initial comparison of Ka- and X-band link performance using an actual deep space communication channel at a relatively modest cost. It is one of the important steps in the continuing evolution of Ka-band capability for deep space communications.

Table 1. Typical Mars Observer Ka-band link parameters

Spacecraft

Output power	20.0 mW
Frequency	33.6 GHz
Antenna diameter	0.5 meter (effective illuminated area)
Antenna efficiency	60.0%
Antenna gain	42.7 dB
Pointing loss	2.71 dB (0.5 degree pointing error)
Carrier Suppression	3 dB at Ka-band

Ground Station

Antenna diameter	34 meters
Antenna gain	79.7 dB
Antenna elevation	30 degrees
Pointing error	0.001 degree
System temperature	29.1 K
Weather	Clear and dry, no wind

Link Reliability

To achieve 90% or 95% reliability, additional margins are required to account for weather and link tolerances. This experiment will record adequate data over a several year time period to define the margins needed for 90% or 95% link reliability.

Carrier Signal-Power-to-Noise-Spectral-Density-Ratio (P_c/N_o)

The objectives of this experiment can be met when the received P_c/N_o ratio exceeds 10 dB. Figure 1 shows P_c/N_o ratios ranging from 60 (18 dB) to 1000 (30 dB); these higher ratios enable increased measurement precision.
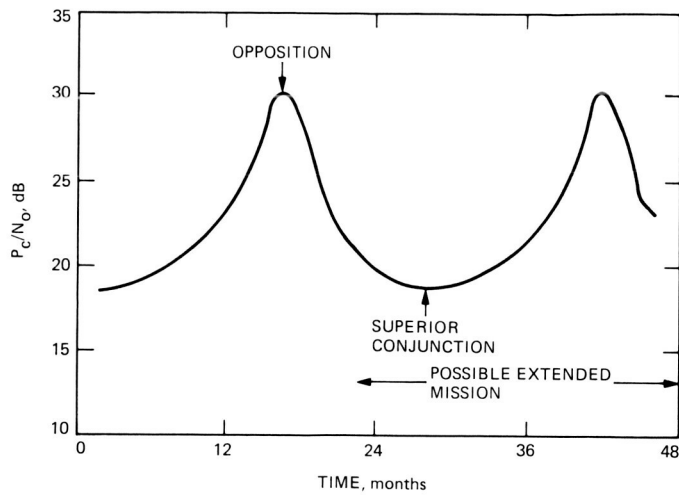
Fig. 1. Ka-band carrier signal-power-to-noise-spectral-density ratio versus versus time expected for 20 mW Mars Observer beacon (see Table 1)
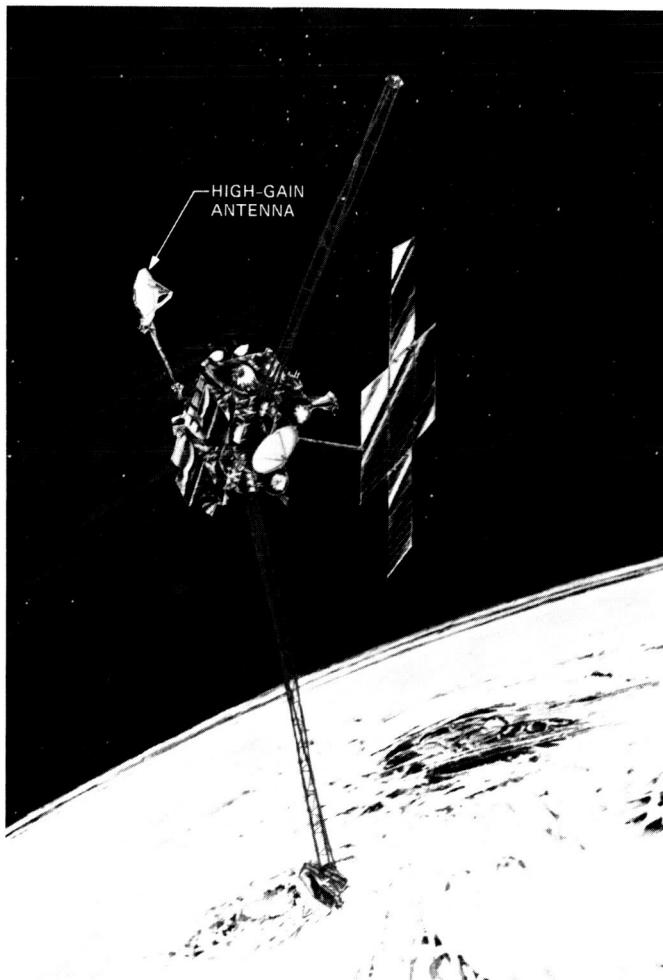
Fig. 2. Mars Observer spacecraft
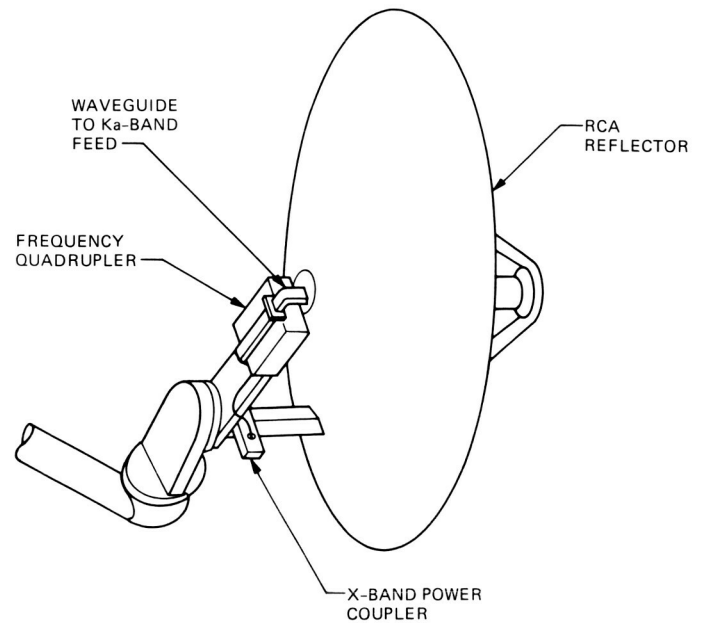
Fig. 3. Mars Observer Ka-Band Link Experiment (KABLE) hardware, back view
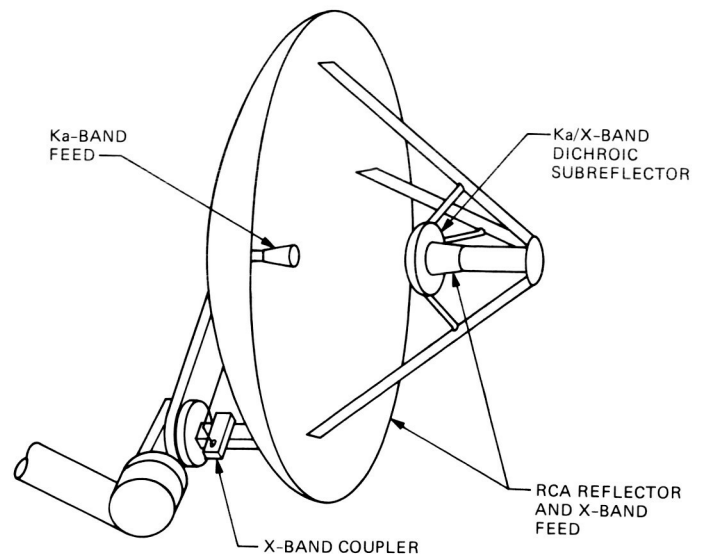
Fig. 4. Mars Observer Ka-Band Link Experiment (KABLE) hardware concept, front view
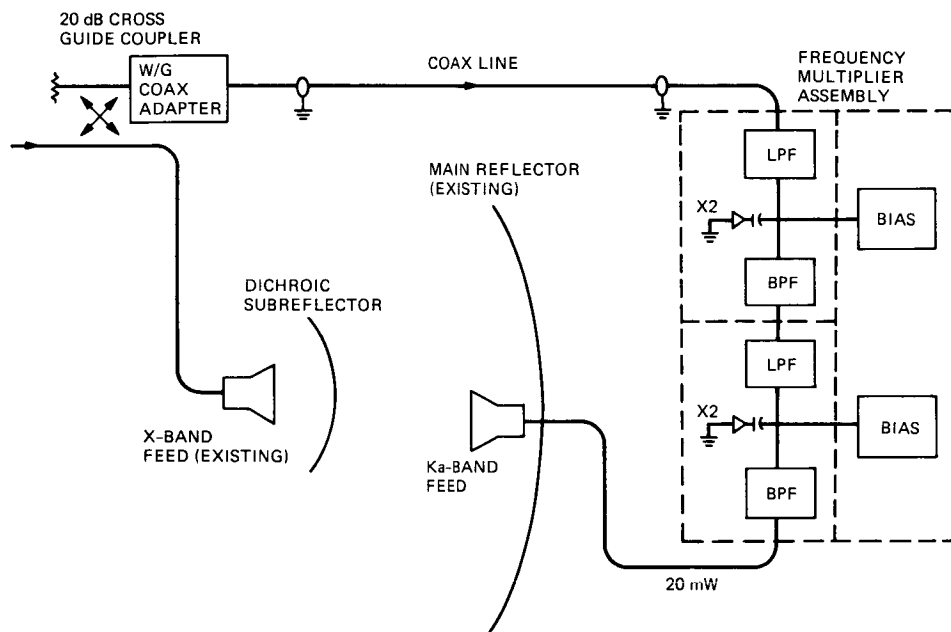
Fig. 5. Mars Observer Ka-Band Link Experiment (KABLE) block diagram

Harmonic Measurements of the Galileo Spacecraft X-Band Transmitter System

P. H. Stanton and F. Manshadi
Radio Frequency and Microwave Subsystems Section

Harmonics of X-band (8.4-GHz) spacecraft transmitter systems can be used to measure the performance characteristics of higher frequency deep space communications links. The Galileo X-band transmitter system was measured at the second, third, and fourth harmonics (16.8 GHz, 25.2 GHz, and 33.6 GHz, respectively). The effective isotropic radiated power was determined to be approximately 52 dBm at 16.8 GHz, 47 dBm at 25.2 GHz, and 25 dBm at 33.6 GHz. These values can be useful, depending upon the sensitivity of the Earth-based receiving system and the distance of the spacecraft from Earth.

I. Introduction

There is a need to accurately measure the performance of the Ka-band (33.6 GHz) deep space communications link between a spacecraft and the Deep Space Network (DSN). This must be done prior to making a commitment to use Ka-band as a primary deep space communications frequency. Harmonics of the current X-band (8.4 GHz) spacecraft transmitter system can be used as "beacons" to enable simultaneous relative measurements of Ka-band and X-band link performance with high differential precision.

The Galileo X-band transmitter system was selected for evaluation at the second, third, and fourth harmonics because a spare transmitter and feed system were available. The objective of the evaluation was to determine the effective radiated power provided by the harmonics of the Galileo X-band transmitter system. During September and October 1985, measurements were made to determine the amplitude and phase patterns of the Galileo X-band antenna feed and to compute the maximum effective radiated power and radiation patterns of the overall Galileo antenna and transmitter system at 16.8 GHz (second harmonic), 25.2 GHz (third harmonic), and 33.6 GHz (fourth harmonic).

II. Measurement of Amplitude and Phase Patterns

The transmission line from the traveling wave tube amplifier (TWTA) output to the antenna feed is of sufficient size to pass higher order modes of the X-band transmitter's harmonics. The transmitted power in each of these higher order modes will be extremely dependent on the configuration of the transmission line. For future use with the actual Galileo spacecraft, the geometrical dependence associated with higher order modes required the measurement of an accurate copy of the Galileo X-band transmission system on an antenna pattern

range. To help compensate for the small dimensional differences between the flight hardware and this test hardware, the operating frequency was swept from 8395 MHz to 8425 MHz to establish a pattern envelope useful in the selection of representative frequencies for the final measurements. The drive power of the TWTA was varied above and below the normal operating power to reflect the differences between TWTAs. The frequencies of 8415 MHz and 8420 MHz and normal operating power were then chosen for detailed measurements.

A test unit support structure was fabricated and the RF equivalent of the Galileo flight X-band hardware, from the TWTA to the antenna feed, was assembled on it. The near-field antenna facility was instrumented and tested as a far-field range for 16.8 GHz, 25.2 GHz, and 33.6 GHz. A schematic diagram of the test chamber setup is shown in Fig. 1. The test unit was mounted on a rotatable platform with the phase center of the feed antenna on the axis of rotation. The patterns were measured by sliding the probe antenna along a track located above the feed. Data were recorded at various angles of feed rotation with increments of 30 degrees. These data were stored on a digital recorder and later corrected for distance variations and test probe patterns. Far-field amplitude patterns were generated, along with the absolute power reference, so that they could be integrated to obtain the total available harmonic power.

The phase patterns of the Galileo X-band feed antenna were needed to complete the data required to compute the effective radiated power of the overall antenna/transmitter system. These phase and amplitude patterns were recorded simultaneously.

III. Calculations of Antenna Radiated Power

The overall antenna effective isotropic radiated power (EIRP) was predicted using the feed pattern measurements. To effectively use the Galileo Ka-band signal for link comparison with X-band, an EIRP of at least 30 dBm is considered necessary for satisfactory reception on Earth. The EIRP is predicted by computation of the overall antenna normalized gain pattern (with respect to the feed) and the feed power measurement. Then the EIRP calculation is modified by the spacecraft antenna reflector mesh loss as well as the subreflector dichroic loss at the frequencies of interest.

The geometry of the Galileo high-gain antenna is shown in Fig. 2. Both the main reflector and the subreflector are shaped surfaces. The feed pattern was measured (as described in previous sections) over 12 azimuthal cuts (phi-cuts) for small but uneven increments in polar angle (theta). These data were first reduced to obtain the feed pattern for every 0.5 degree

in theta direction over all the phi-cuts. Then the Cassegrainian reflector system was replaced with an equivalent paraboloid (Ref. 1). This yields a good approximation for the overall antenna pattern while saving considerable time and computer costs. For the equivalent paraboloid, the dish diameter is the same as the main reflector diameter (D = 188.8 in. or 479.55 cm) and the effective focal distance was calculated to be F_e = 195.22 in. or 495.86 cm. The far-field pattern of the paraboloid was computed using a modified version of the physical optics/Jacobi-Bessel polynomial expansion program.[1] This program was developed for circularly symmetric feed patterns; however, for this case, the feed pattern for harmonic frequencies is not circularly symmetric. Therefore, the PATRN subroutine of Rahmat-Samii (Footnote 1) was modified to generate the feed pattern, at any point required by the program, via a two-dimensional interpolation of the feed measured data. Using this program, the far-field patterns of the Galileo antenna for second, third, and fourth harmonics of 8415 and 8420 MHz were computed. The phi = 0 cut patterns for the above frequencies are shown in Figs. 3 through 8. These patterns represent the normalized antenna gain relative to the feed gain, i.e., they are computed by assuming zero gain for the feed boresight. It can be seen that the higher the frequency is, the more lobes are generated near the boresight. This is due to the interaction of a larger number of propagating modes at higher frequencies.

To find the feed radiated power, its RF pattern was measured relative to a known reference antenna. This was accomplished by moving a probe in front of the reference and test (Galileo feed) antennas. In this measurement, the far-field patterns of both antennas, as well as the input power of the reference horn, were found. Taking into account the distance between the probe and the reference and test antennas, the overall EIRP can be determined. The results are shown in Table 1. Considering the 30-dB minimum EIRP requirement, it is evident that the fourth harmonic output of the antenna would not be useful at a distance greater than 1 AU, but the second and third harmonic signals should provide more than enough power for accurate measurements.

IV. Losses

The reflector mesh of the Galileo high-gain antenna is identical to the tracking and data relay satellite (TDRSS) antenna built by the Harris Corporation. For the TDRSS antenna, the peak gain loss (compared to a solid reflector),

[1]Rahmat-Samii, Y., *Offset Parabolic Reflector Computer Program for Analysis of Satellite Communication Antennas*, JPL document D-1203, Jet Propulsion Laboratory, Pasadena, CA, Dec. 1983 (JPL internal document).

as a function of frequency, is computed and shown in Fig. 9.[2] It can be seen that the mesh loss for the second and third harmonics of the Galileo transmitter (16.8 GHz and 25.2 GHz) are approximately 0.5 dB and 1.22 dB, respectively; these losses would not reduce the EIRP of the antenna significantly.

Test samples of dichroic subreflector material were measured in the laboratory and found to have reflection loss of between 3 and 6 dB, between 18 and 40 GHz. Phase shift as a function of incident angle was not measured. Making a gross simplification by ignoring the phase, the dichroic reflected power loss is included in the final EIRP given in Table 1.

V. Conclusions

The second, third, and fourth harmonic effective radiated power of Galileo's X-band transmission system was computed based on measurements of the transmitter and feed system without main or subreflector. The results of this study indicate useful EIRPs at the second and third harmonics. The fourth harmonic EIRP is too low to be considered useful at the distance planned for the Galileo mission. The results are sufficiently encouraging to consider this approach for other missions such as Magellan, where the distance from Earth will vary from 0.3 AU to 1.7 AU.

Acknowledgment

References

1. Hanan, P. W., "Microwave Antennas Derived from the Cassegrain Telescope," *IEEE Trans. Antennas and Propagation*, Vol. AP-9, pp. 140–153, March 1961.

2. Agrawal, V. D., and Imbriale, W. A., "Design of a Dichroic Cassegrain Subreflector," *IEEE Trans. Antennas and Propagation*, Vol. AP-27, No. 4, pp. 466–473, July 1979.

Table 1. Effective isotropic radiated power (EIRP)

Frequency	EIRP w/o Mesh or Subreflector Loss, dB	Mesh and Subreflector Loss, dB	EIRP w/Mesh and Subreflector Loss, dB
33.660	32.0	5.1 to 8.1	26.9 to 23.9
33.680	32.0	5.1 to 8.1	26.9 to 23.9
25.245	54.4	4.2 to 7.2	50.2 to 47.2
25.260	52.0	4.2 to 7.2	47.8 to 44.8
16.830	57.7	3.5 to 6.5	54.2 to 51.2
16.840	56.7	3.5 to 6.5	53.2 to 50.2
8415	96.3	0	96.3
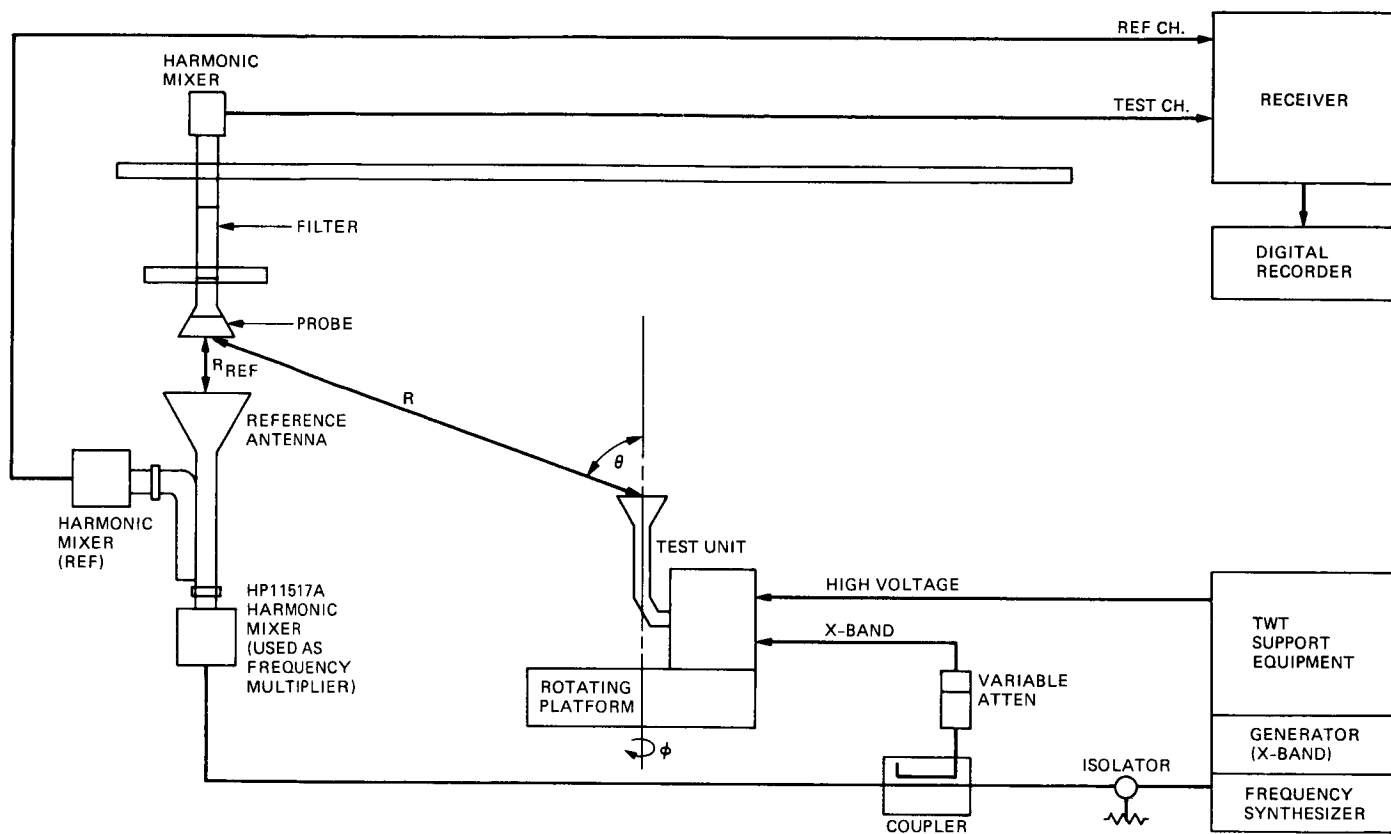
Fig. 1. Near-field facility setup for far-field radiation pattern measurements

S-BAND FOCUS

X-BAND PHASE CENTER
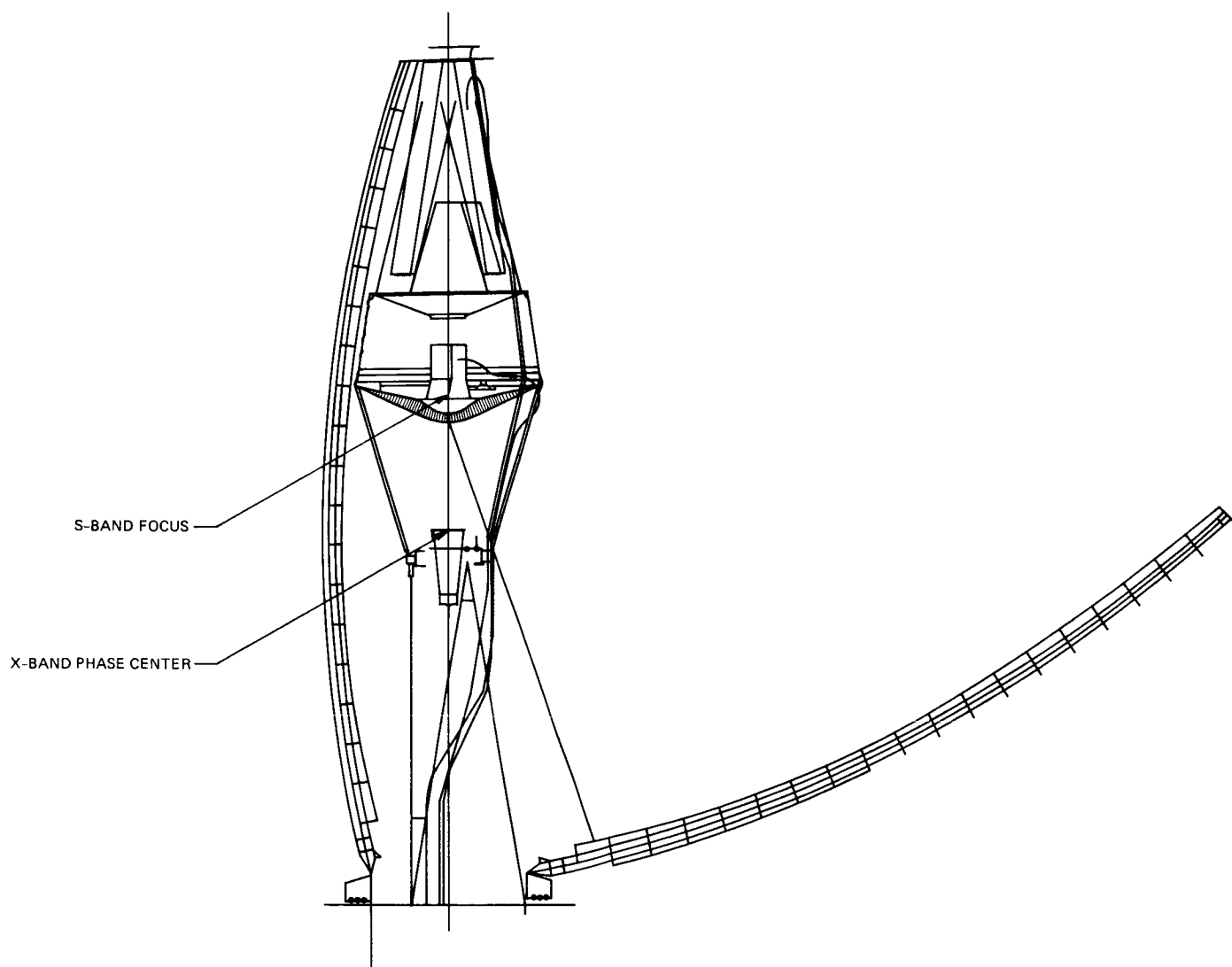
Fig. 2. Galileo HGA configuration
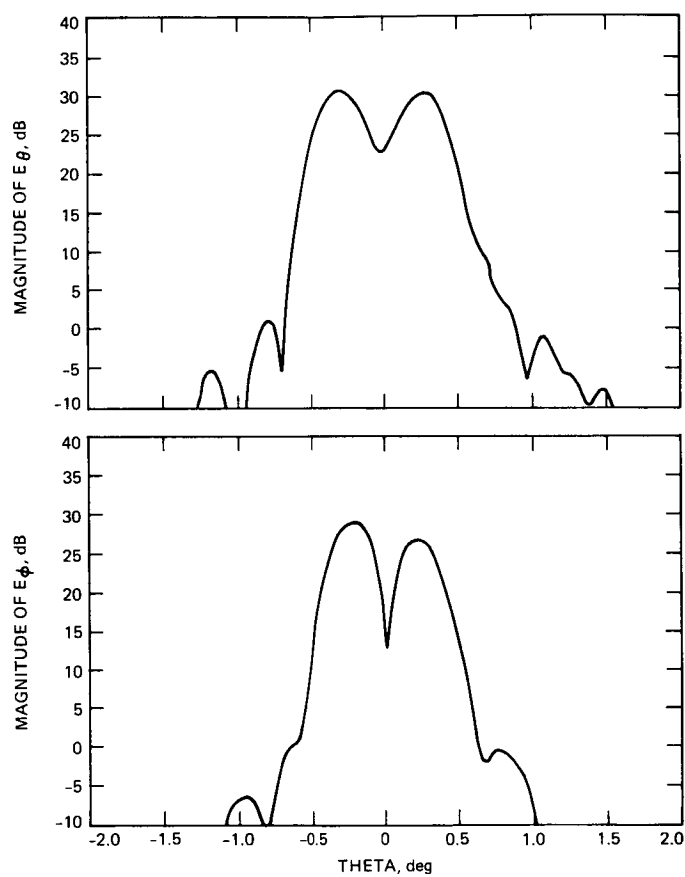
Fig. 3. Far-field patterns for $\phi = 0$ degree plane at second harmonic of 8415 MHz
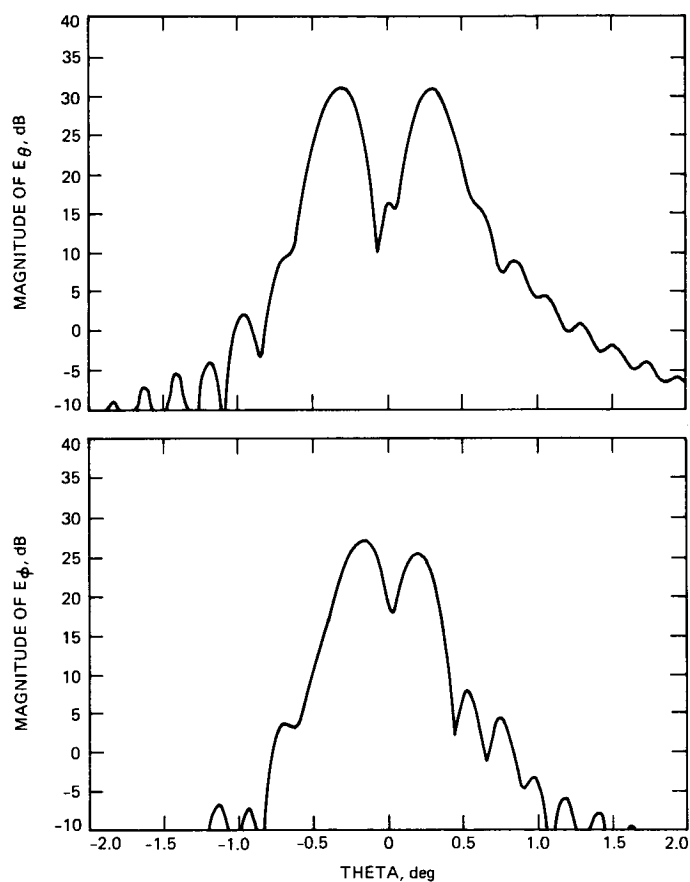
Fig. 4. Far-field patterns for $\phi = 0$ degree plane at second harmonic of 8420 MHz
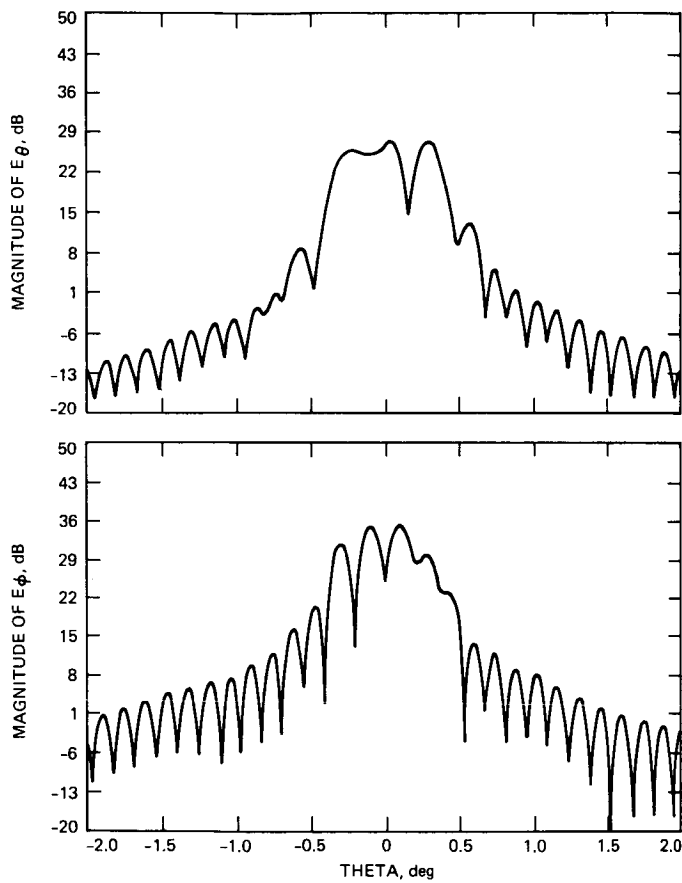
Fig. 5. Far-field patterns for $\phi = 0$ degree plane at third harmonic of 8415 MHz
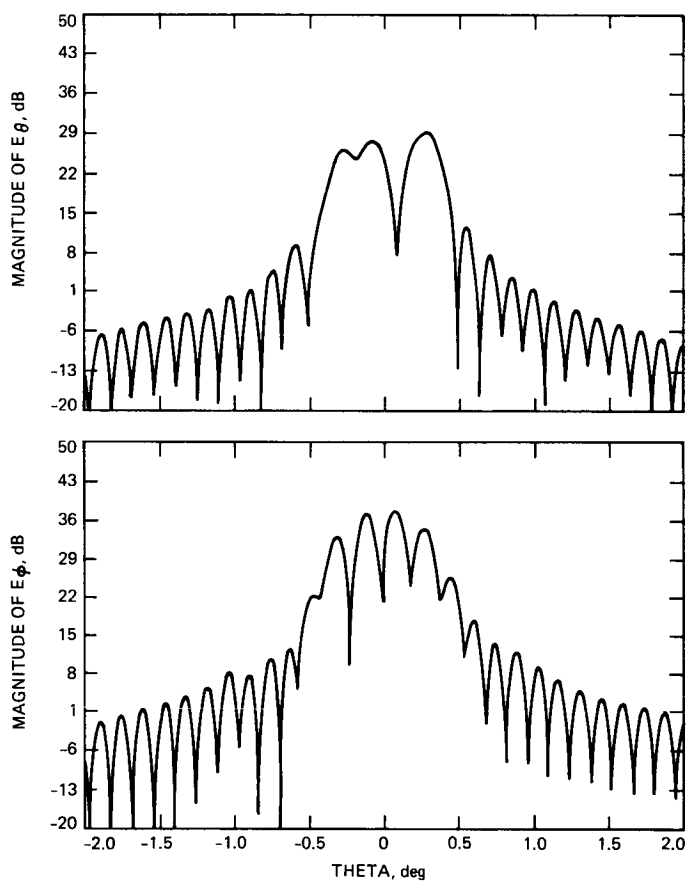
Fig. 6. Far-field patterns for $\phi = 0$ degree plane at third harmonic of 8420 MHz
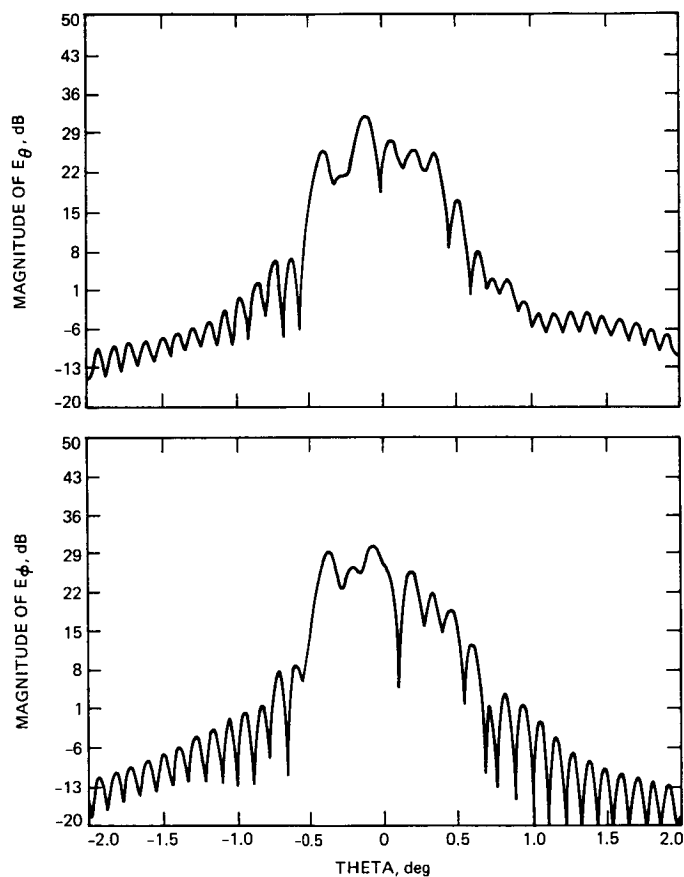
Fig. 7. Far-field patterns for φ = 0 degree plane at fourth harmonic of 8415 MHz
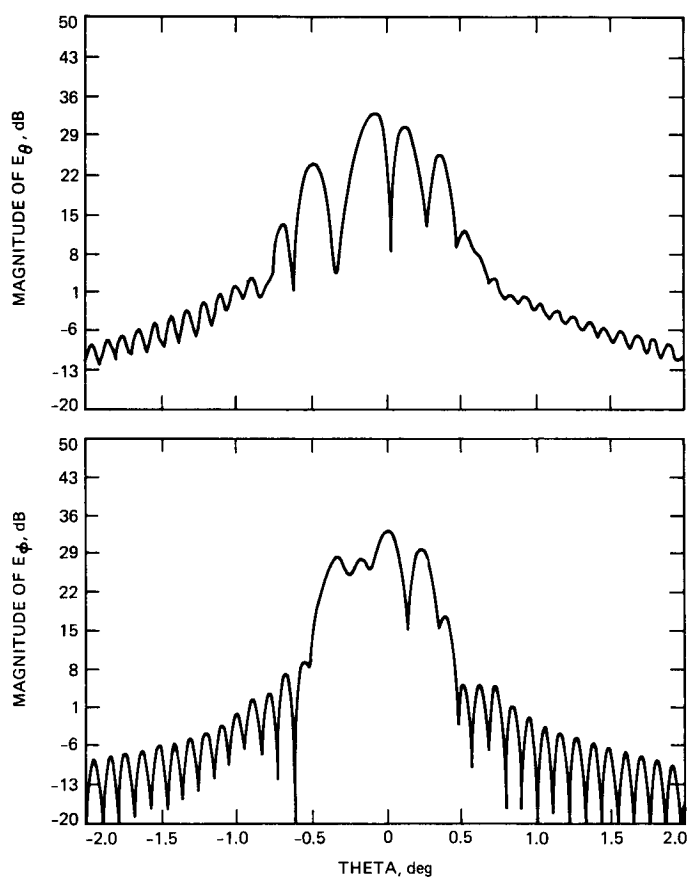
Fig. 8. Far-field patterns for φ = 0 degree plane at fourth harmonic of 8420 MHz
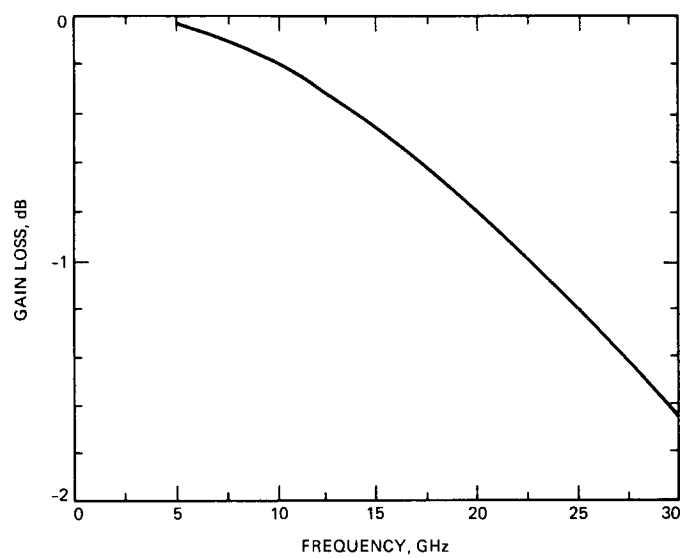
**Fig. 9. Boresight gain loss vs frequency for TRDSS mesh
parameters**

Proposed Upgrade of the Deep Space Network Research and Development Station

Joel G. Smith
TDA Technology Development Office

Continued exploration of the solar system will require continued evolution of capabilities to support deep space communication and navigation. That evolution will rely, as it has in the past, on the development, demonstration, and field testing of new communication and navigation technologies. The existing Deep Space Network (DSN) research and development station, DSS 13, at the Venus site, Goldstone, California, has been instrumental in those prior developments.

However, the present antenna is no longer able to provide the necessary support for new technologies. The 26-meter antenna has good performance at S-band, fair performance at X-band, but is unusable at the anticipated Ka-band (32 GHz). It is not suitable for conversion to beam waveguides, and is not usable as a test bed for demonstrating high efficiency because of structural pliancy (it's too floppy). Additionally, its size and age are increasingly a liability in demonstrations.

A new 34-meter beam waveguide version of the existing DSN high efficiency (HEF) antennas has been proposed for the FY88 Construction of Facilities (C of F) budget. The antenna is to be built at the Venus site, adjacent to the old antenna, and serve as the DSN research and development antenna through the end of the century.

I. Introduction

The growth of the deep space communication and navigation capability that has occurred since inception of the solar system exploration program in the late 1950s has been in response to the increasing demands of that program for more data per second over greater distances. That evolution has come through a variety of improvements: bigger spacecraft antennas, more spacecraft power, higher operating frequencies, better means of modulation and coding, and better ground facilities. These improvements have followed an orderly transition: concept, analysis, laboratory models, field tests, demonstrations, and commitment to mission use.

Fundamental to that process, both for improved spacecraft and ground capability, has been a research and development station within the DSN for testing and demonstrations. The major such site for the last two decades has been the Venus site at Goldstone, California, with its 26-meter antenna, DSS 13.

Research and development have been conducted at other locations in the past within the DSN, but the lower priority of non-flight project status and heavy loading of the network has caused research and development to suffer when attempted on operational stations.

A separate research and development facility is most cost-effective, and that facility must be of sufficient quality that the environment of the test bed parallels the applications environment, and the capability does not mask the required observations.

II. Past Contributions From the Venus Site

The list of new capabilities demonstrated and/or field tested at the Venus station is restricted here to those which are still visible today in the network and which were technically significant achievements.

One of the earliest test beds at Venus was an S-band planetary radar system used for probing Venus (hence, the station name). This verified the "emptiness" of space at S-band and demonstrated wideband (MHz) ranging, ranging codes, and transmission and reception on the same antenna. Additional work at the Venus Site included the unified S-band concept of shared telemetry and tracking on one signal, computer-controlled subsystems, digital on-line signal processing using computers, wideband correlation (a precursor to VLBI), programmable up-and-down links, precision power monitoring, and digitally controlled spectral analysis.

Other examples are low-noise dual mode feeds, dual simultaneous S-band transmissions proposed for Viking commanding (tests showed that intermodulation effects precluded simultaneous transmissions of two narrowly spaced S-band carriers), verification of analytic models for the design of DSS 14 (the Venus antenna was subjected to a "shake" test, and the computed and measured resonant frequency values were compared and found to agree), and verification of telemetry arraying capability with demonstration for Mariner Venus Mercury 1973.

III. Present Contributions From the Venus Site

The Venus station could, until recently, under control from the Jet Propulsion Laboratory, acquire, track, command, and receive telemetry from deep space spacecraft operating on S-band (commanding was limited to the Pioneer 8 spacecraft). These demonstrations (from 1978 through 1984) of the ability to remotely turn on, tune up, control, and turn off a station

were instrumental in the decision to develop remote controlled station operation in the DSN.

The X-band uplink system now planned for implementation in the network is based on a prototype built at the Venus station and undergoing tests since 1982. The 20 kW system is remotely controlled and phase stable to 10^{-15} over 10^3 seconds, a good intermediate goal in the effort to stabilize the DSN stations to 10^{-16}.

The common aperture feed now on the Venus station is a new design now used on the 34-meter HEF antennas, and is being considered for implementation on the 70-meter antennas. It replaces the dichroic plate and small reflector now used. It will allow simultaneous dual transmission (both S- and X-band, up to 20 kW) and reception (both S- and X-band at 25 Kelvin noise temperature). A next generation feed capable of the same performance, but at 100 kW levels, is feasible.

A 7 km optical fiber cable, carrying six separate fibers, with multi-GHz bandwidth capability, was buried roughly five feet below grade from the Echo site to the Venus site. This promises to provide 10^{-17} stable time and frequency standard transfer between stations. The benefits are reduced numbers of Hydrogen maser frequency standards, and the possibility of lower cost navigation by means of connected element interferometry.

IV. Present Antenna Characteristics

The present research and development station at the Venus site (see Fig. 1 for layout) has many key capabilities used in the past and useful in the future. These include: 1) best available radio frequency isolation from operational stations (on the order of 200 dB); 2) suitable control buildings, electronics laboratories, and collimation tower; 3) buried optical fiber link to the Echo site; 4) super-power transmitter test facilities (400 kW); and 5) high bay for feed cone testing.

However, the present research and development station does not have one critical capability needed for the future: a large precision antenna. The present antenna (see Fig. 2) lacks several key characteristics:

(1) It is too pliant (floppy). It has good S-band performance and fair X-band performance, but it cannot be upgraded to serve as a Ka-band test bed. It cannot be converted to serve as a beam waveguide test bed, and it cannot be used as a test bed for improving antenna efficiency.

(2) It is too small. All deep space antennas have been enlarged to 34-meter size or larger. The 26-meter

antenna is too small for cost-effective transfer of technology.

(3) It is too old. The antenna will be 33 years old at planned tear down in 1992. It has been "band-aided" frequently to extend its life. Recently, outrigger beams were added to compensate for a cracked, irreparable foundation to allow DSS 13 to be used through 1992.

V. Proposed New Antenna Characteristics

The antenna selected to replace the present one is a 34-meter diameter design of demonstrated high performance at X-band, with tolerances of 0.025 inch. In fact, it is planned to use the recently constructed 34-meter HEF antennas (Fig. 3) as a model, but modified to incorporate beam waveguide.

Other possible antenna designs were considered. Offset clear aperture antennas were rejected because: 1) the small potential gain-to-temperature (G/T) benefit over a center-fed antenna with improved subreflector and subreflector support; 2) the added cost associated with the non-symmetric construction; and 3) the lack of transfer of that technology to any of the existing DSN antennas.

A stiffer antenna, with tolerances of 0.015 inch, and high performance at Ka-band, was rejected in favor of the more pliant design because: 1) the stiffer antenna cost several million dollars more than the pliant; 2) the pliant antenna with 30 to 50 percent efficient performance at Ka-band can be upgraded to 70 to 80 percent efficient performance by electronic means; and 3) the upgrade of the pliant antennas is a technology that is transferrable to existing DSN antennas.

A smaller antenna was rejected in favor of the 34-meter design because: 1) scaling smaller antenna performance at higher frequency (for example, with a 17-meter antenna at 64 GHz) is not possible due to media properties and receiving equipment unique to deep space communication bands; 2) transfer of 70-meter microwave and pointing is possible over the 2:1 size ratio; and 3) transfer of 34-meter microwave, pointing, and structural technology is straightforward.

The specific beam waveguide design includes either a conventional method (Fig. 4) or a unique "bypass" method (Fig. 5) because, the bypass method allows the 34-meter antenna design to use the same structure as the present HEF antennas; but it does not appear to compromise performance.

The antenna proposal calls for the antenna to be completed under the FY88 C of F budget, up to but not including the subreflector and subreflector supports. These items are to be developed under DSN Advanced Systems Program funding to provide maximum impact on antenna performance.

The antenna will utilize either an alidade room or pedestal room inside the wheel-and-track to house the electronics. The room envisioned routes the microwave beam received and/or transmitted through multiple fixed mirrors, and movable mirrors to one of several work stations. RF isolation will permit research and development to be done at work stations switched out of the beam waveguide system while transmission/reception occurs at a work station switched in.

VI. Planned Research and Development

The following research and development tasks have been identified so far for the Venus station when the new antenna becomes operable in early 1990.

(1) Demonstration of subreflector and subreflector supports with high antenna efficiencies at X- and Ka-bands.

(2) Demonstration of low threshold acquisition and tracking of spacecraft telemetry.

(3) Demonstration of fast acquisition of X-band spacecraft telemetry.

(4) Demonstration of precision blind pointing at X- and Ka-bands.

(5) Demonstration of comparative reception performance at X- and Ka-band under various weather conditions.

(6) Demonstration of simultaneous dual frequency reception, at both X- and Ka-bands, and tracking of various spacecraft experimental beacons.

(7) Demonstration of improved reception performance of a beam waveguide configuration at X-band in the rain.

(8) Demonstration of the improved reliability, maintainability, operability, and repairability of transmitters and low-noise receivers in a beam waveguide environment.

(9) Demonstration of simultaneous transmission of high power (400 kW) and low-noise reception (−180 dBm) at X-band through beam waveguides.

(10) Demonstration of cryo-cooled multi-element Ka-band array feeds and low-noise amplifiers for higher antenna efficiency.

(11) Demonstration of ultra-stable (10^{-16} to 10^{-17}) station frequency stability.

(12) Demonstration of Ka-band radar.

(13) Demonstration of improved operator monitor and control interfaces.

VII. Conclusion

The new research and development antenna planned to be operable in early 1990 is timely because of the aged state of the present "band-aided" antenna, and because of the needed development and demonstration of Ka-band capability in the DSN in time for use by the Cassini mission in the mid-1990s, and subsequent deep space missions.

This antenna promises to allow the Venus site station to continue its tradition of field testing new capabilities for the operational DSN with a tool sufficient to meet the needs through the end of the century.
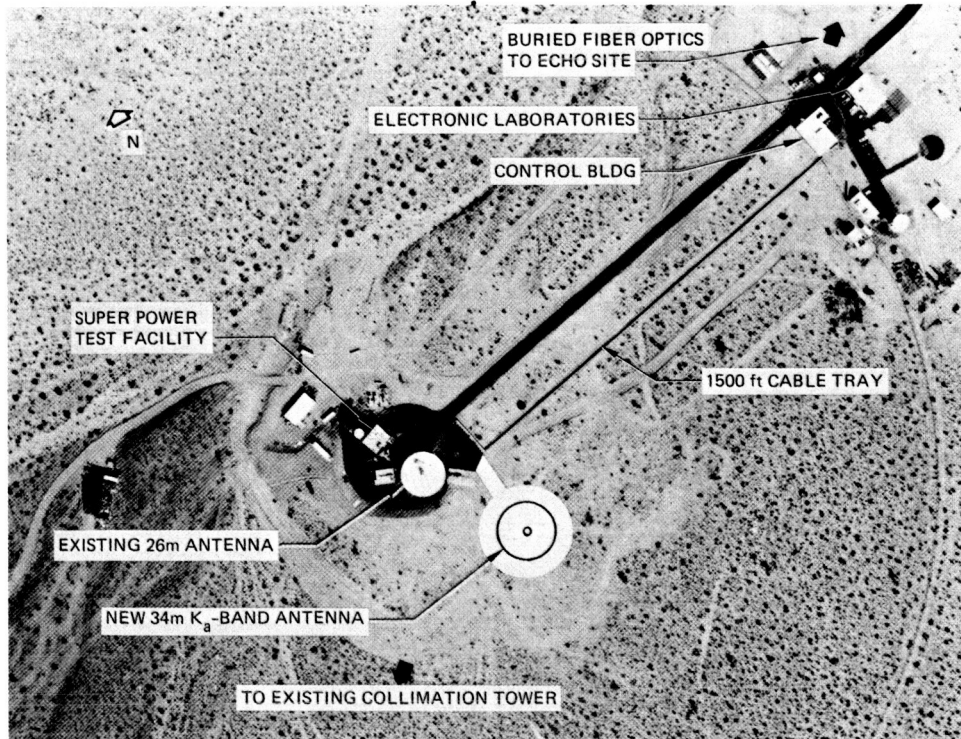
Fig. 1. Layout of R&D station
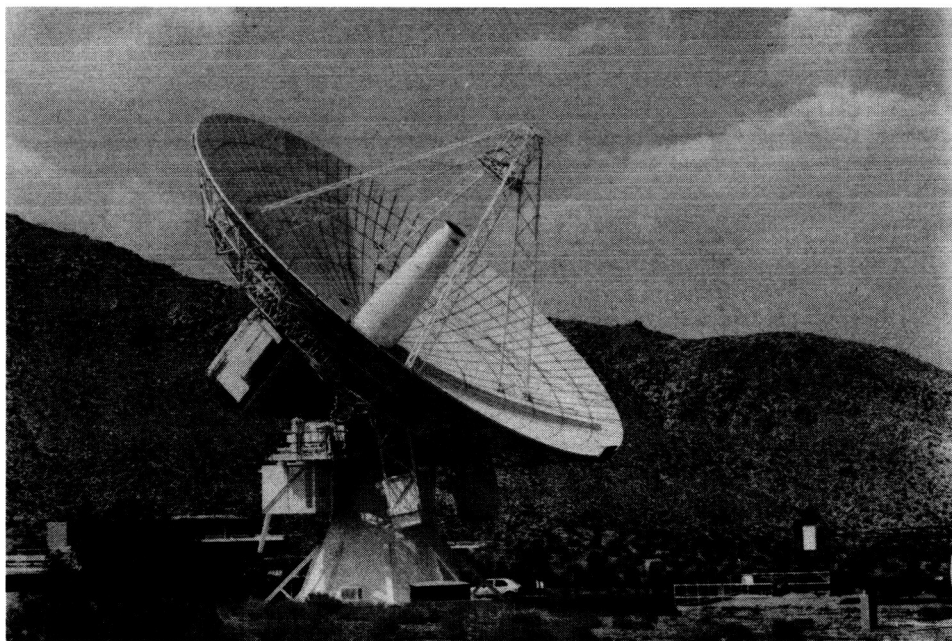
Fig. 2. Present R&D antenna
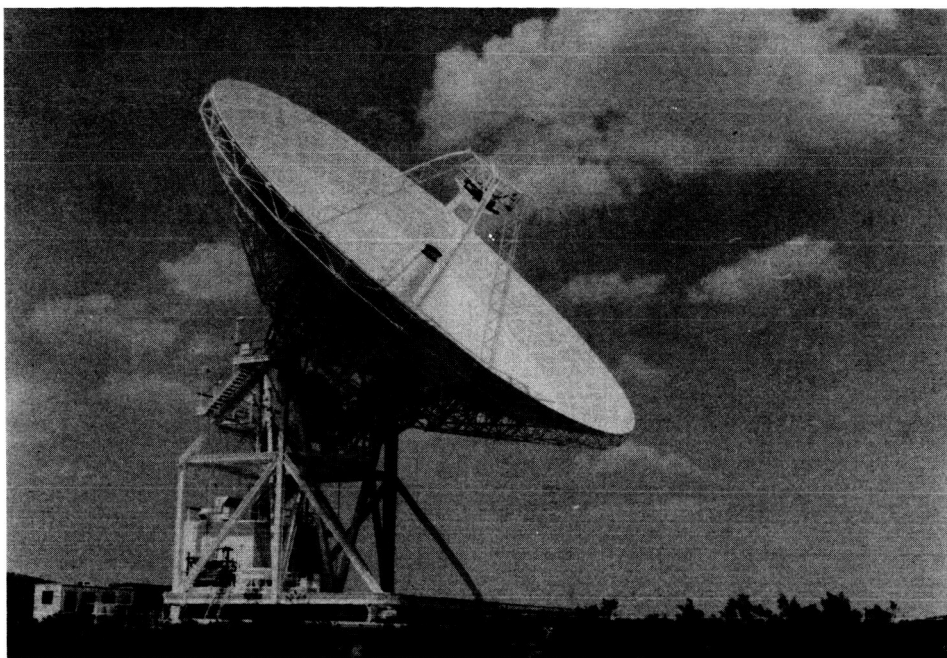
Fig. 3. Proposed R&D antenna
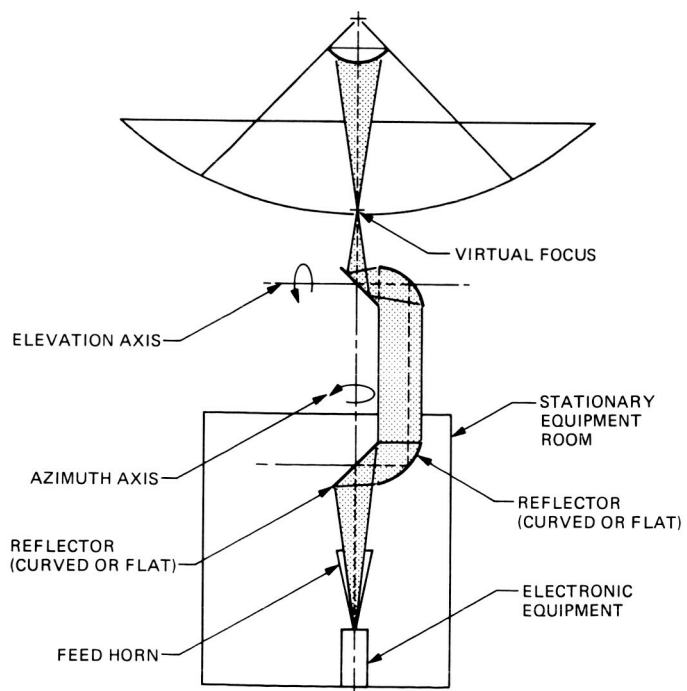
Fig. 4. Conventional beam waveguide
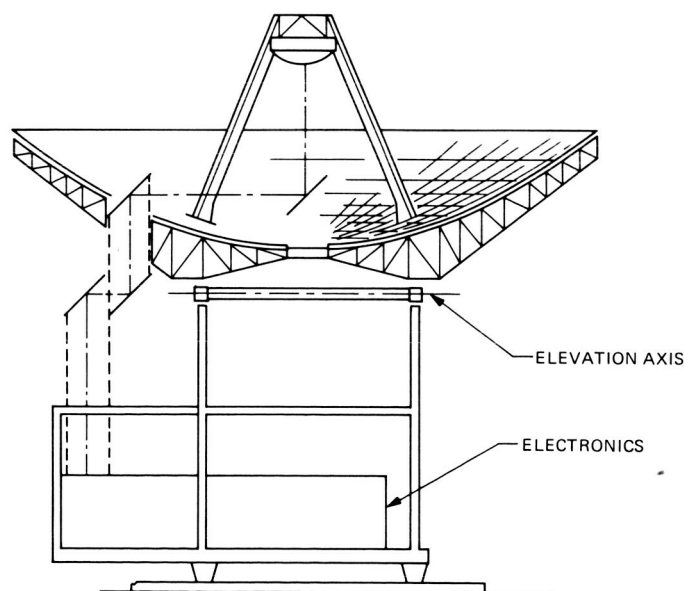
Fig. 5. Lateral bypass beam waveguide

163

Ka-Band (32 GHz) Spacecraft Development Plan

A. L. Riley
Spacecraft Telecommunications Equipment Section

A road map for the development of a protoflight 32 GHz spacecraft solid state transmitter is given. The major milestones include the development of device and component technology required for use in the spaceborne experimental and operational transmitter systems. Two experimental spacecraft transmitter systems are envisioned: first, a low power beacon, to determine the performance of a 32 GHz downlink communications system; and second, an array feed, to further verify the results of the first experiment and serve as a test bed for technology required for an operational system. The first experiment has been proposed to NASA Headquarters for flight aboard the Mars Observer spacecraft with spacecraft integration in early 1989. The second is to be available for integration aboard a spacecraft such as the Comet Rendezvous Asteroid Flyby (CRAF) mission in the 1990 time frame. These experimental systems are to lead to the development of a protoflight transmitter for subsequent spacecraft integration in 1992, the time frame of the Cassini mission to Saturn.

I. Introduction

A fair amount of study has been dedicated to the application of spaceborne solid state transmitter systems to 32 GHz deep space communications. This solid-state technology was chosen because most of the missions studied to date appear to benefit most from it. Electron beam devices such as the traveling wave tube amplifier tend to be more attractive for higher power applications. Furthermore, solid-state active arrays give the added benefit of a simple method to compensate for antenna pointing errors due to spacecraft attitude errors by utilizing electronic beam steering (EBS). Hence, the discussions here emphasize solid state technology.

II. Solid-State Technology State-of-the-Art

Gallium Arsenide (GaAs) field effect transistor (FET) devices are just becoming available for moderate power appli-

cations at 32 GHz. Laboratory devices operating with 100 mW of output power with 30% power added efficiency have been demonstrated by the Texas Instruments Central Research Laboratory and others. Sample devices with 1.5 W of output power with 15% efficiency are anticipated to be available in the last quarter of 1986 from Japanese vendors.

Figure 1 is a plot of the output power performance for several devices demonstrated by Texas Instruments and by the General Electric Research Laboratories. These data include performance of both MESFET and MODFET, or HEMT, devices. Typical performance for a 32 GHz HEMT is shown in Fig. 2.

Power-added efficiency is a key element of cost savings for the 32 GHz transmitter. Overall cost of spacecraft power as a function of amplifier system efficiency is shown in Fig. 3 for various levels of transmitter RF power. System efficiency

includes DC power supply efficiency (assumed here to be 85%) and RF to regulated DC power conversion efficiency. It assumes that the cost per Watt of RTG power is $250K. The cost of power is inversely proportional to efficiency with a cost savings of approximately $6.5 M for an 8 W transmitter if the efficiency is increased from 15% to 30%. Such an increase in efficiency results in a more than $1.5 M savings for a 2 W transmitter.

The technology development needs required to implement a high efficiency 32 GHz spaceborne transmitter system focus on FET devices, monolithic phase shifters, array combining, and EBS. The FET devices need improved power handling capability (100–250 mW), improved power added efficiency (greater than 30%), and high reliability (passivation, greater quantities). The monolithic phase shifters need higher frequency of operation and lower loss; the array combining area will concentrate on integrated microwave and feed circuits; EBS requires development of algorithms and software.

III. Development Approach

Figure 4 lays out a road map for development of a spacecraft transmitter for 32 GHz operation in the Cassini mission time frame. The four elements are the development of transmitter devices, a Ka-band link experiment (KABLE) with the Mars Observer mission, a Ka-band communications experiment (KACE) with the CRAF mission, and a prototype transmitter for the Cassini mission.

The development of high efficiency 32 GHz FET amplifiers and phase shifters will exploit rapidly evolving technology development programs presently funded by DoD and NASA, take advantage of the recently announced $1 billion (FY 87–91) DoD Microwave/Millimeter Wave Monolithic Integrated Circuit (MIMIC) Program, cooperate with Lewis Research Center in order to focus part of its present Texas Instruments (TI) efforts on JPL needs, and direct JPL funding of FET development to meet specific needs.

The proposed KABLE experiment consists of a transmitter aboard the Mars Observer spacecraft providing a Ka-band downlink to a DSN receiving station. The transmitter consists of a passive frequency multiplier driven by an X-band signal tapped from the power amplifier output to produce a signal at the fourth harmonic (approximately 33.7 GHz) of the X-band downlink. This Ka-band signal is fed to the Mars Observer high gain antenna for transmission to Earth. From the spacecraft transmitter development perspective, this experiment will provide initial experience with Ka-band onboard a deep space spacecraft at very low cost. KABLE is described in detail in an companion article (Ref. 1).

IV. KACE

A block diagram of the CRAF Ka-band Communications Experiment (KACE) is shown in Fig. 5. The KACE is a downlink beacon system supplementing the prime X-band link. The KACE hardware consists of a Ka-band exciter, a Ka-band amplifier module, and a feedhorn and dichroic subreflector to diplex Ka- and X-band signals. The exciter and amplifier module hardware will be developed for use on CRAF and other Ka-band missions. In later missions, a larger number of these modules will be assembled in an array feed to produce higher power levels.

The amplifier module is one of the more technically challenging aspects of the system. This module is required to have high power-added efficiency (greater than 25%), small size and low cost. One approach to meeting these requirements is to use GaAs monolithic microwave integrated circuit (MMIC) FET devices. The development of devices similar to those required for the KACE system is presently being carried out by TI and funded by Lewis Research Center. Discussions are presently underway to focus this effort toward the development of a device to meet the KACE requirements. Figure 6 is a conceptual drawing of the device and layout of the monolithic circuit consisting of four FET devices on a single GaAs chip. The output stage consists of two parallel FETs driven by two series stages. It is not clear at this point whether a total monolithic structure will yield the greatest efficiency since no tuning of the critical output stage can be done. An alternate approach is to use a discrete output stage mounted in the same package as the input monolithic stages. Several industrial laboratories are presently developing discrete devices. The antenna system will be composed of the baseline Viking antenna with an X-band feed at the focal point. Modifications include the addition of a frequency selective subreflector illuminated by a Ka-band horn feed. A new feed support strut will be required for CRAF, with or without KACE, so that a negligible cost increase will result if the strut is designed to support the subreflector.

Calculations of a Ka-band link with KACE have been made. These assume 0.25 W of amplifier output, 0.2 dB circuit loss, and a Viking antenna with 48.4 dB gain and 0.25 degree pointing error (resulting in 2 dB loss). A typical set of link parameters is given in Table 1 for a 70-meter DSN antenna at a 25 degree elevation angle. A 40 bps data rate can be maintained at Goldstone to 17 AU with 95% link reliability. Assuming the same values of transmitter power, circuit loss, and antenna characteristics, but with different stations, data rate, and range, the performance shown in Table 2 would be maintained.

The KACE will require about 3 W of spacecraft power and could be operated intermittently or continuously during the

cruise phase of the mission. The experiment will measure link performance and acquire link reliability statistical data over varying elevation angles and weather conditions. The mass of the experiment hardware will be of the order of 5 kg or less, and, including the subreflector, the electronics will occupy a volume less than 2,000 cubic cm.

V. Cassini Strawman Spacecraft Ka-Band Design

The Cassini Ka-band system will be composed of some new components and those demonstrated in the CRAF KACE. The KACE components include the exciter and amplifier modules combined in an active array. New features of the Cassini 32 GHz system include multiple modules, the addition of phase shifters and new radiating elements with each module, and beam steering electronics to compensate for spacecraft pointing errors.

A strawman Cassini 32 GHz downlink system developed by Boreham[1], and shown in Fig. 7, consists of an active array feed system illuminating a 3.66 meter high gain antenna by way of a Cassegrainian subreflector. The subreflector is a frequency selective design to allow illumination of the high gain antenna by X- and S-band feeds at the focal point. The X-band feed is used for communications uplink and back-up downlink, and the S-band is used for a radio science experiment. One promising design that was investigated consisted of a 21 element array with five elements across the array and an element spacing of 1.8 wavelength. Since the half-power beam angle of the 3.66-meter reflector at 32 GHz is 0.18 degrees, and the Mariner Mark II baseline attitude pointing accuracy is 0.11 degrees, with a resulting scan loss of as much as 4 dB, either greater spacecraft stability or beam steering was required.

The strawman design assumes EBS by the array controlled by attitude sensors, which reduces the scan loss to 0.5 dB. The spaceborne 32 GHz system consists of an exciter similar to the type developed for the KACE with the active array. Major elements of the active array are a low loss power distribution system, a set of phase shifters, amplifier modules and radiating elements and the beam steering control system. The amplifier is similar to those developed for KACE. These modules produce output power levels of 0.1 W so the 21 element

[1]Boreham, J. F., "A 21 Element EBS Array Feed for the SOTP Spacecraft," JPL IOM 3360-86-030 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 22, 1985.

EBS array feed power amplifier (AFPA) radiates 2.1 W of RF power into a 3.66-meter HGA and supports the same data rate as a previously proposed X-band baseline system. That X-band system consisted of redundant 10.6 Watt solid-state amplifiers (SSAs). The Ka-band and X-band designs were assumed to have the same fixed mounted HGAs and use the same attitude control accuracy of 0.05 degree. It was assumed that the attitude control system provides the only X-band beam pointing, but the Ka-band beam is provided with additional fine pointing with EBS. The attitude control system provides the Ka-band EBS controller with deadband position information to accomplish the vernier pointing. Details of the control system are shown in Fig. 8.

Figure 9 is a schematic diagram of the Ka-band EBS AFPA. The assembly consists of 21 ports, redundant DC to DC power converters, and redundant EBS controllers. The legend below the block diagram in Fig. 8 shows the RF power levels at the array and FET module inputs to and the outputs for the Ka-band configuration. The Ka-band exciters will use power output modules similar to those in the AFPA modules. A 4.0 dB transmission loss from the exciters (through the hybrid) and to the AFPA input and the power divider loss sets the input power levels shown in Fig. 9. The output levels from the modules are set by the required total radiated power divided by the number of elements in the array.

Figure 10 shows details of the FET module designs, which include an input phase shifter, a monolithic gain stage, a discrete driver stage, a final high efficiency power stage, and an output isolator. It has been assumed that efficiencies of discrete devices in the 30 to 100 mW output level range will equal or exceed 30% in production units by 1991. (This assumption is reasonable by the technology cut-off year for a 1995 Cassini launch based on the present rate of technology advances.) Assuming a DC to DC converter efficiency of 85%, and 1.5 W for the EBS controller, results in a total of 11.4 W for the 21 element array. Mass estimates for the EBS array feed portions of the Ka-band system are shown in Table 3.

VI. Conclusion

A development plan has been laid out to prepare the solar system exploration program to exploit Ka-band for downlink operation by the mid 1990's, in time for the Cassini mission. The plan takes advantage of development by other agencies where that is possible, with adoption of those developments for deep space applications. Flight experiments are considered as well as a protoflight model.

References

1. Riley, A. L., et al., "A Ka-band Link Experiment (KABLE) with the Mars Observer," *TDA Progress Report 42-88*, Jet Propulsion Laboratory, Pasadena, California, February 1987.

Table 1. Telecommunications system downlink performance estimate

Parameter	Design Value
Transmitting System	
RF power output, dBm	24.0
Transmitter circuit loss, dB	−0.2
Antenna Gain (referenced to CP), dB	48.4
Antenna ellipticity, dB	1.5
Antenna pointing loss, dB	−2.0
Path Parameters	
Space loss (where range = 17.5 AU and frequency = 31.9 GHz), dB	−310.9
Receiving System	
Antenna gain for matched polarization, dB	84.8
Antenna ellipticity, dB	1.0
Antenna pointing loss, dB	−0.2
Noise spectral density, dBm/Hz	−183.3
System noise temperature, K	33.7
Zenith noise temperature, K	23.2
Additive noise for elevation angle, K	10.5
Elevation angle, deg	25
Telemetry Performance Estimate	
Required pt/no. (ranging off), dB Hz	23.5
Data rate, bps	40.0
Required E_B/N_0, dB	4.2
Required carrier margin, dB	10.4
Modulation level (RMS), deg	43.3
Performance margin, dB	3.8
Link reliability	1.0
Sigma, dB	0.7

Assumptions:

Ka-band/0.25 W KSSPA/HGA circuit
Ka-band/1.47 M body-fixed HGA/0.25 deg pointing error
DSN 70 M station/Ka-band/mechanical compensation/PE = 0.001 deg.
Goldstone/extrapolation from Ka-band noise temperature measurements/worst
DSN Block III receiver/10.8 Hz bandwidth mode
Telemetry channel/Viterbi (K=7,R=1/2,Q=3), 40 bps, PB=1.E-3

Table 2. KACE telemetry performance estimates

Station	Maximum Range, AU	Data Rate, bps
Goldstone	17.4	40
Goldstone	5.6	768
Canberra	13.7	40
Canberra	5.6	454

Table 3. Mass estimates of Cassini Ka-band EBS array feed

Components	Mass/Unit, g	Number of Units/Assembly	Mass, g
FET Module	12	21	250
Phase Shifter	9	21	190
Isolator	5	21	105
Array Element	5	21	105
Power Splitter	150	1	150
Power Converter	950	2	1900
EBS Controller	600	2	1200
Structure	250	1	250
Wiring	100	1	100
Hardware	100	1	100
Total Mass			4350
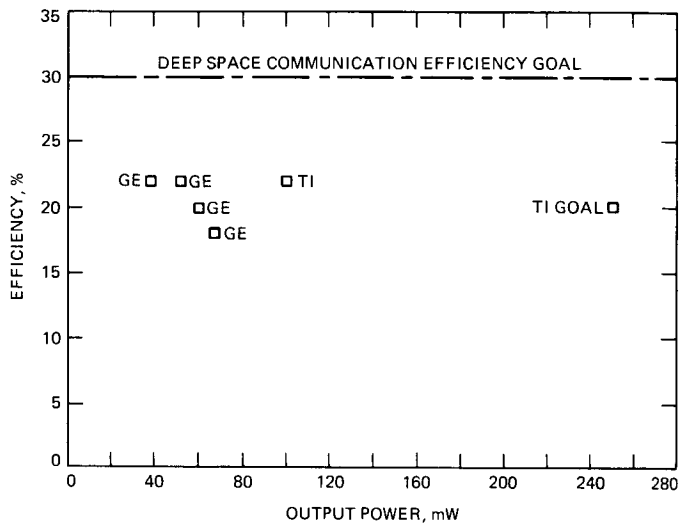
Fig. 1. State-of-the-art FET performance, 32 GHz
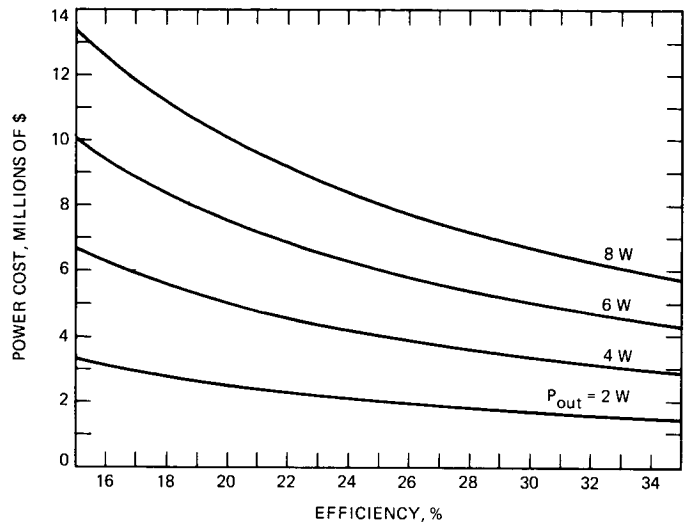
Fig. 3. PA power cost vs efficiency
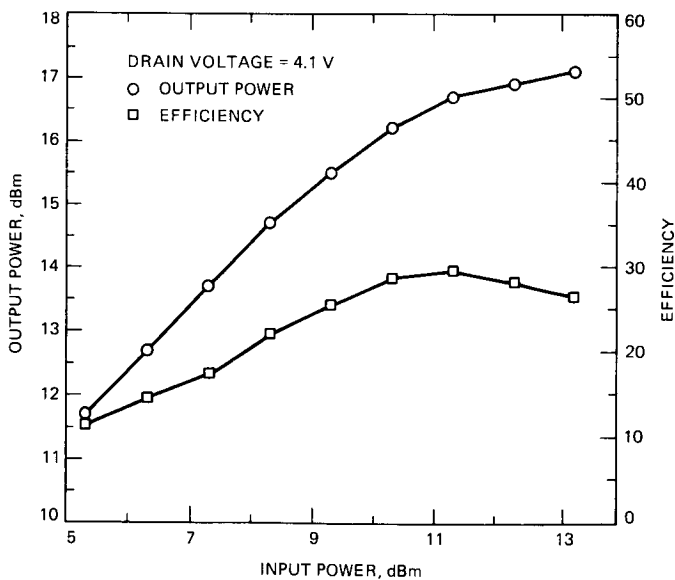
Fig. 2. Gain compression of a 32.5 GHz HEMT
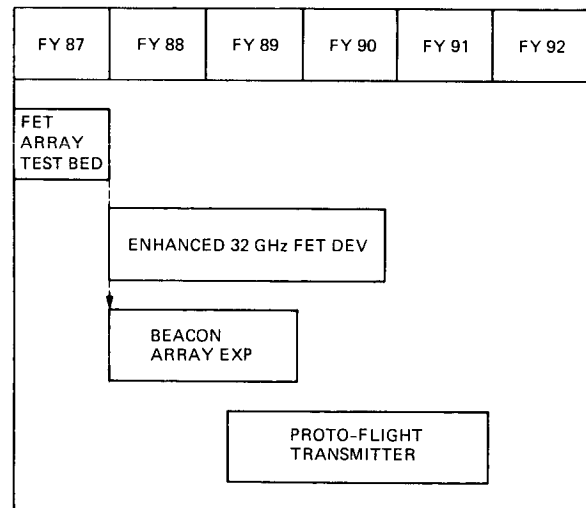
Fig. 4. Development of the 32 GHz space transmitter
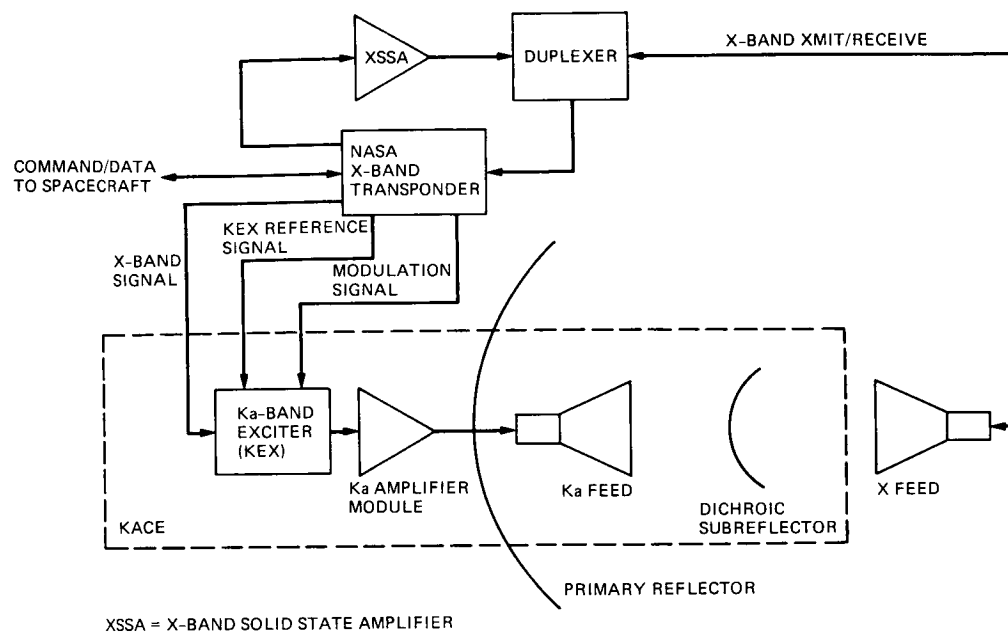
Fig. 5. Interfaces for the CRAF Ka-band communications experiment (KACE)
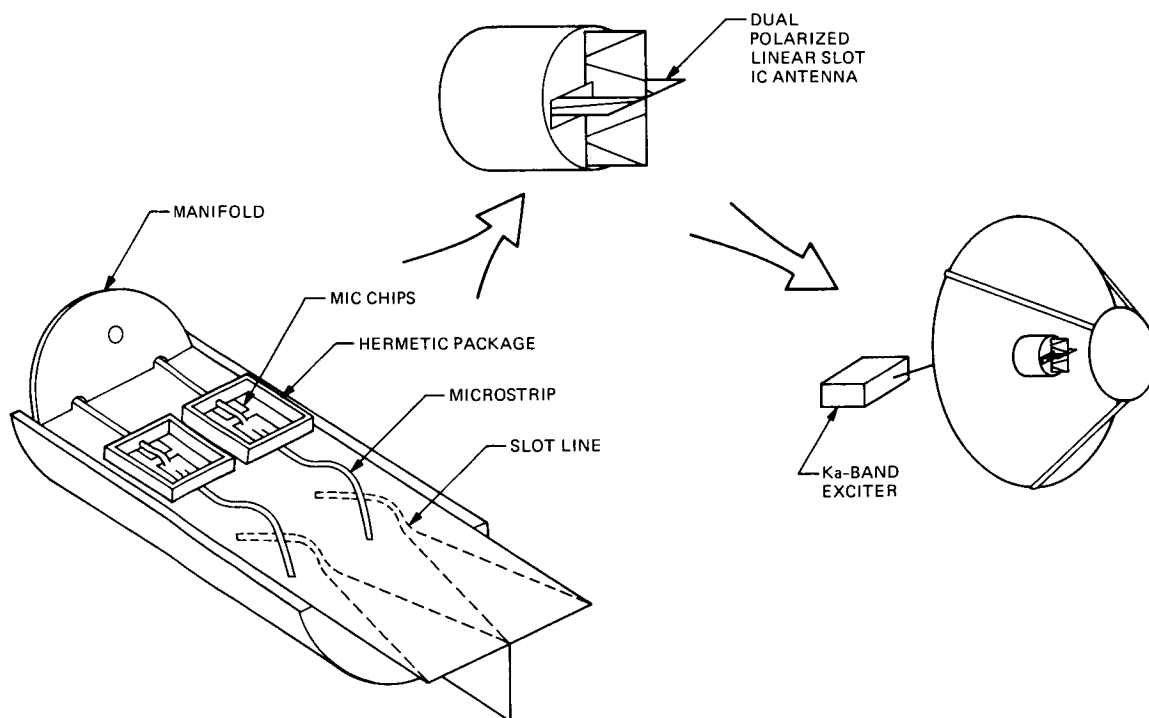
Fig. 6. Configuration of the Ka-band beacon experiment system
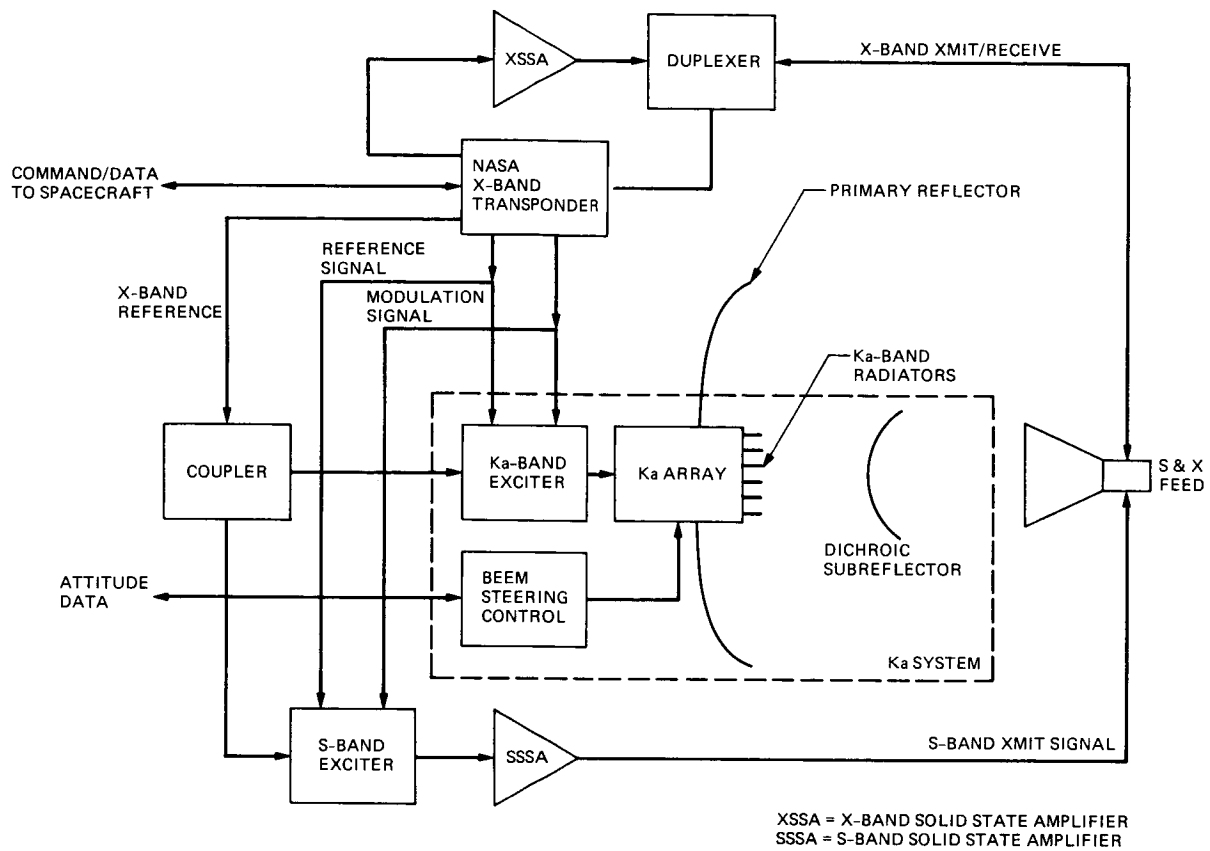
Fig. 7. Cassini Ka-, X-, and S-band communications system
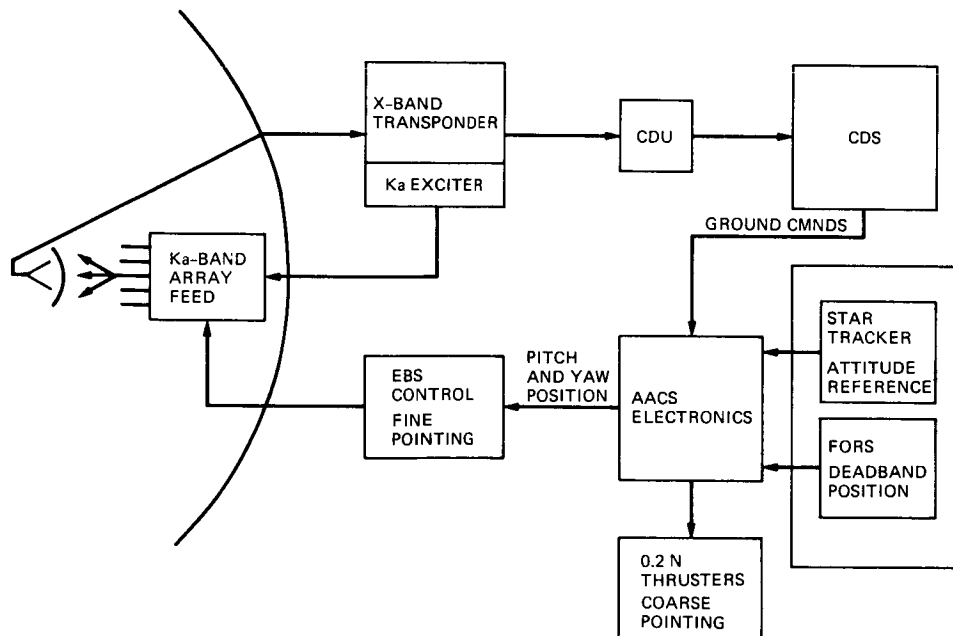
Fig. 8. Functional block diagram of the EBS antenna pointing system

21 ELEMENT EBS AFPA	TOTAL P_{in}	MODULE P_{in}	MODULE P_{out}	TOTAL P_{out}
RF POWER dBm	+17	+2	+20	+33.2
mW	50	1.6	100	2.1 W
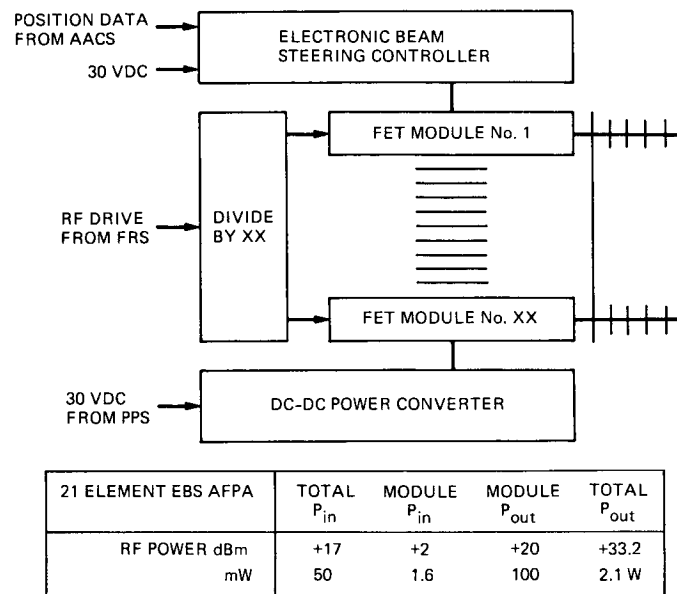
Fig. 9. Block diagram of the Ka-band EBS AFPA
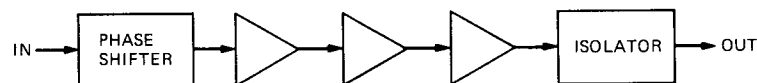
21 ELEMENT FET MODULE, 3.66 m HGA

DEVICE TYPE	PHASE SHIFTER	MONOLITHIC GAIN STAGE	DISCRETE DRIVER STAGE	DISCRETE POWER AMP	ISOLATOR	TOTALS
POWER LEVEL, dBm	+2.0	0	+10.0	+15.3	+20.3	+20.0
GAIN/LOSS, dB	-2.0	+10	+5.3	+5.0	-0.3	+18.0
EFFICIENCY, %	-	15	25	30	-	25.0
DC POWER, mW	-	60	95	245	-	400

Fig. 10. The 21-element Ka-band AFPA FET module design
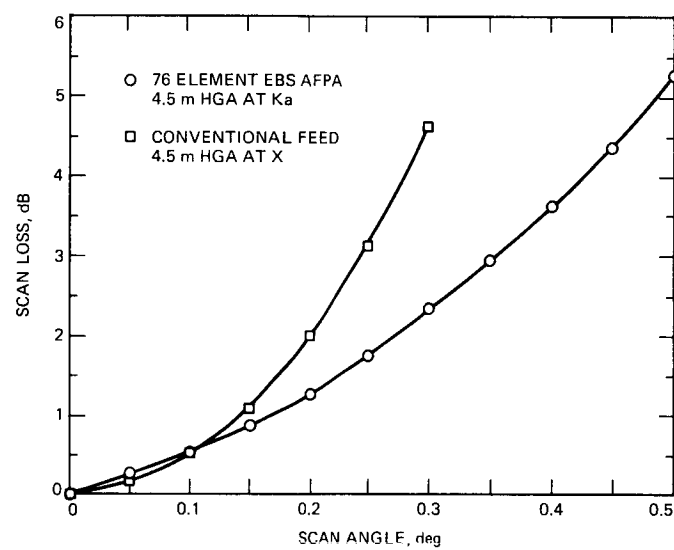
Fig. 11. The 4.5-m HGA scan/pointing losses for EBS AFPA and conventional feeds

Beam Waveguides in the Deep Space Network

R. C. Clauss and J. G. Smith

Telecommunications and Data Acquisition Technology Development Office

A beam waveguide is a mechanism for guiding electromagnetic radiation from one part of an antenna to another through a series of reflectors. Appropriate placement of reflectors on an antenna allows a "beam" to be guided "around the elevation axis" and/or "below the alidade." The beam waveguide permits placement of all electronics in a room on the alidade below the elevation axis, or below the alidade; feed horn covers to be protected from the weather; and feed electronics to be in spacious rooms rather than in crowded cones, and always level rather than tipping with change in elevation angle. These factors can lead to lower costs in new implementation such as Ka-band, better antenna performance at X-band, more efficient and stable performance of transmitters and receivers, and lower maintenance and operating costs. Studies are underway to determine methods for converting the major antennas of the Deep Space Network (DSN) to beam waveguide operation by 1995.

I. Introduction

A beam waveguide is an open waveguide where electromagnetic propagation does not depend on the boundary conditions of the walls. A beam waveguide system typically consists of an arrangement of reflectors that direct the beam from one location to another. The set of reflectors may be contained in large tubes, where minimal interaction occurs with the walls of the tubes.

Beam waveguide systems are used on antennas as a means of moving the focal point to a more convenient location. For example, on the Weilheim, Germany, 30-meter antenna shown in Figs. 1 and 2, the secondary focal point is located behind the main reflector, beyond the elevation bearing on the elevation axis. This has the primary advantage of allowing space for a large room that contains a dual frequency feed horn, two

X-band masers, two S-band parametric amplifiers, receivers, and a transmitter; it also allows the electronics to remain level and keeps the feed horn cover dry despite rain.

Beam waveguide systems typically use curved reflectors or combinations of curved and flat reflectors. Flat reflectors simply redirect the beam; curved reflectors may refocus the beam in order to control beam divergence along the waveguide. Curved reflectors may be ellipsoidal, or paraboloidal; the shapes are sometimes modified to favor one frequency. Design considerations are addressed elsewhere (Ref. 1).

Beam waveguide systems can use a pair of reflectors along the elevation axis to bring the beam down to a feed electronics room on the alidade. Again, the reflectors can be flat or curved as desired. A second pair of reflectors can be used to advantage

along the vertical axis to bring the beam to a pedestal room that is below the alidade (Ref. 2). Figure 3 shows these two configurations conceptually; Fig. 4 shows one of the three COMSAT antennas at Roaring Creek, Pennsylvania, where all feeds and electronics are within the stationary buildings on which the three 32-meter antennas are mounted. Figures 5 and 6 show the 64-meter deep space antenna at Usuda, Japan, with its three-story alidade building. The spacious interior of the Usuda alidade rooms provide a laboratory environment for transmitters, cryogenically cooled low-noise amplifiers, receivers, and much other equipment (see Fig. 7).

Beam waveguide systems can include many reflectors, including movable ones, in order to move the beam to any of several desirable locations. The 45-meter radio astronomy antenna at Nobeyama, Japan (Fig. 8), has the configuration of reflectors shown in Figs. 9 and 10. Note that the reflectors can be moved to redirect the beam. Vernier pointing of the antenna beam can be accomplished by movement of reflectors. Figures 11 and 12 show such a system for an aircraft-mounted 98-GHz radiometer, using a 43-cm × 63-cm scanning reflector (Ref. 3).

The DSN has used reflectors in a fashion similar to those in beam waveguides for over 10 years. As Fig. 13 shows, the dual-frequency (S-band and X-band) arrangement of dichroic plate and ellipsoidal reflector is like a beam waveguide system. However, these do not move the beam below the elevation axis and/or alidade. Hence, today DSN antennas contain their front-end electronics in cones and rooms that are crowded and that tip.

It is proposed to retrofit beam waveguide systems into existing DSN antennas, and to employ them on all new DSN antennas. These beam waveguides will be used to direct the beam below the elevation axis and/or the alidade, to provide vernier pointing where needed, to provide frequency sensitive beam splitting, and to redirect the beam with a movable reflector to allow the use of different or redundant sets of electronic equipment.

II. Performance Gains for the DSN

All advantages to the DSN will accrue because sophisticated electronic equipment now mounted above the elevation axis in DSN antennas without beam waveguides can be mounted in rooms below the elevation axis once beam waveguides are added. This means that the traditional single or multiple cones and transmitter rooms above the elevation axis can be eliminated.

Instead, all electronics will be housed in either a stationary feed room that does not tip or rotate (typical of COMSAT), or in a feed room that rotates, but does not tip (typical of Usuda and Nobeyama).

This has many implications for improved performance and operation. First, the crowded cones are replaced with spacious rooms in which new implementations can be installed without expensive rework. The electronics in the antenna are accessible all of the time — no need for downtime for maintenance on the electronics. Equipment can be moved in or out with no loss of antenna time, and without need for use of the antenna structure itself as a crane for moving heavy equipment.

Second, equipment can be simpler and easier to maintain within operating conditions. Cryogenic system gas lines will no longer traverse hundreds of feet from alidade platforms around the elevation axis. The lines will be short and fixed with no flexible sections. The long bundles of power and signal cables now running from alidade to above the elevation axis will largely disappear. Cables will be short and fixed. There will be no gas lines, water lines, or cables above the elevation axis except for electrical connections for the subreflector control and aircraft warning lights.

Transmitters will be cheaper to build and easier to maintain. For example, a non-tipping transmitter can use steam rather than water cooling — in fact, an increase in output power is possible without changing water flow rate because of increased efficiency of cooling through steam. Tolerances required for the proposed 34-GHz gyroklystron will be easier to obtain with a stationary design than with one designed to be tilted.

Cryogenic systems will also be cheaper and more reliable. Stationary units that store large volumes of liquid helium can be used to operate at temperatures below 4.2 K and extend mean time between failure for cryogenic systems from 2000 hours to 20,000 hours. Operating at temperatures near 1.5 K will reduce maser noise temperature by a factor of three, triple maser gain-bandwidth product, and allow operation of superconducting frequency standards in the front-end area equipment room on the antenna.

Frequency stability will be improved by eliminating flexing cables, controlling the environment for critical equipment and cable, and measuring and stabilizing or compensating in real time for beam waveguide length variations.

Finally, beam waveguides can provide increased antenna performance at X-band in the rain. This is because systems can be built so that feed horn covers and dichroic plates are not exposed to moisture. Light rain that should cause only 2 K additional noise at X-band can add 40 K due to water in the feed horn cover and dichroic plate with our current system.

The resulting signal-to-noise-ratio loss of 4 dB can be circumvented with beam waveguides. Figure 14 shows relative performance during rain at Weilheim, Germany, and as specified in the DSN at the 95% weather probability.

III. Converting the DSN

Tests conducted with a DSN S-band traveling wave maser on the 64-meter antenna in Usuda, Japan (Ref. 4) showed that the zenith system temperature was 1 K lower than that of a DSN 64-meter antenna, while at 30 degree and 10 degree elevation angles, the Japanese system temperature was 2.3 K and 3 K lower than the DSN's (Fig. 15). Usuda's lower noise temperature is thought to be due to less microwave scattering off the smaller quadripod support of the Usuda antenna.

There does not appear to be any major cost difference between an antenna with or without beam waveguide when building and equipping a new antenna. Probably the beam waveguide system has a lower cost. That is, the cost of the reflector system and large feed equipment room is more than offset by the reduced costs from not providing a cone, transmitter room, heavier quadripod, added counterweights, added cryogenic lines, added cabling, and the added cost of "tight" layout design in the crowded space above the elevation axis.

Thus, since neither lower performance nor higher cost should arise, it appears to be appropriate to build only beam waveguide designs for future DSN antennas. The new research and development antenna planned for the Venus site will not only utilize beam waveguides, but also be used to conduct extensive tests to answer questions about how the DSN can best utilize beam waveguide. Replacements for the old HA-DEC 34-meter antennas are expected to use beam waveguides. The remaining issue deals with conversion of the existing 64-/70-meter and 34-meter high efficiency (HEF) antennas.

Figure 16 shows one "conventional" beam waveguide configuration that is suitable for a 34-meter antenna (Ref. 5). For the HEF antennas, this would require rework of the elevation axis truss — that is, replacement of the horizontal beam with a "donut" support to allow the beam to be directed through to the center of the dish. This configuration is a possibility for the replacements for the old HA-DEC antennas. Figure 17 shows a "bypass" mode that transfers the beam "through" the reflector, but requires no structural changes in the elevation axis truss. This is a candidate for the HEF antenna since the horizontal support can be retained; both configurations are being considered for tests in the new research and development antenna at the Venus site.

Beam waveguide conversion for the 70-meter antennas remains uncertain. Initial surveys show that at least one approach can place the beam within the central column now used for the master equatorial pointing system. This could permit the pedestal area to be used for all transmitter and receiver equipment now housed above the elevation axis.

This assumes, however, that the 70-meter antennas will not use a master equatorial pointing system. Studies are now underway to find ways of converting to beam waveguide without loss of the master equatorial pointing system, and to find ways of pointing the 70-meter antenna without the master equatorial system.

IV. Concluding Remarks

The advantages of beam waveguides to DSN future implementation and operation are significant. Long range plans call for conversion of the DSN antennas during the 1990s. Replacement of the old HA-DEC 34-meter antennas is slated for the early 90s, conversion of the 70-meter antenna for the mid-90s, consistent with plans to add Ka-band reception; conversion of the 34-meter HEF antennas are less definite, but expected to occur in the late 90s.

References

1. Veruttipong, T., Withington, J., Galindo-Israel, V., Imbriale, W., and Bathker, D., "Design Considerations for the Beam-Waveguide Retrofit of a Ground Antenna Station," *TDA Progress Report 42-87*, pp. 10–23, Jet Propulsion Laboratory, Pasadena, California, November 15, 1986.

2. *Satellite Communications Engineering*, Miya, K., editor, Lattice Company, Ltd., Tokyo, Japan, 1975.

3. Wilson, W. J., Howard, R. J., Ibbott, A. C., Parks, G. S., and Ricketts, W. B., "Millimeter-Wave Imaging Sensor," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-34, No. 10, pp. 1026–1035, October 1986.

4. Fanelli, N. A., Goodwin, J. P., Petty, S. M., Hayashi, T., Nishimura, T., and Takano, T., "Utilization of the Usuda Deep Space Center for the United States International Cometary Explorer (ICE)," *Proceedings of the Fifteenth International Symposium on Space Technology and Science*, Tokyo, Japan, 1986.

5. Katow, S., Cucchissi, J. J., Chuang, K. L., Lansing, F. L., Levy, R., Menninger, F. J., and Stoller, F. W., "Structural Design Options for the New 34-meter Waveguide Antenna," *TDA Progress Report 42-88*, this issue, Jet Propulsion Laboratory, Pasadena, California, February 15, 1987.
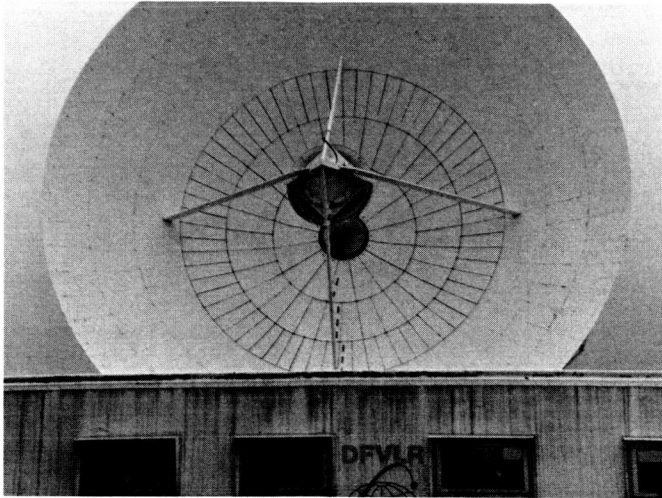
Fig. 1. Weilheim 30-m antenna viewed from front
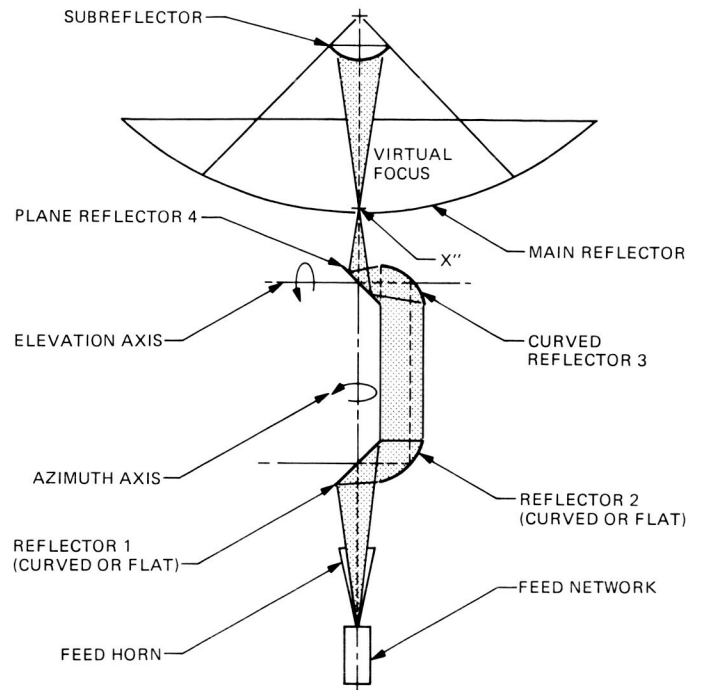
Fig. 3. Conceptual configuration of a beam waveguide
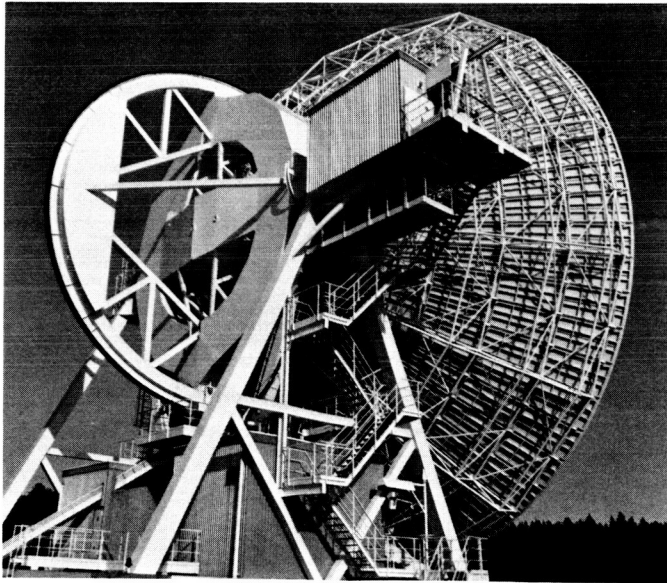
Fig. 2. Weilheim 30-m antenna viewed from rear
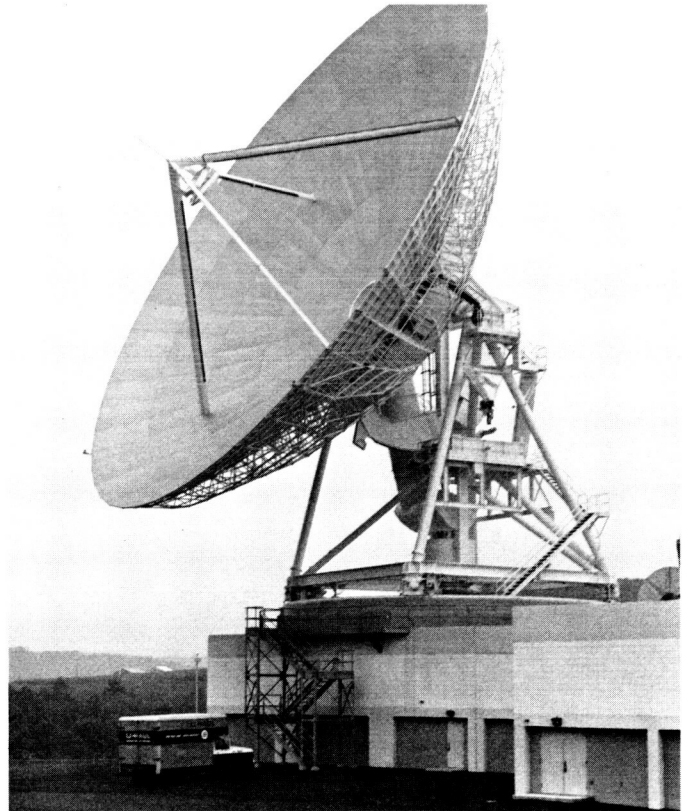
Fig. 4. 32-m COMSAT antenna at Roaring Creek, Pennsylvania
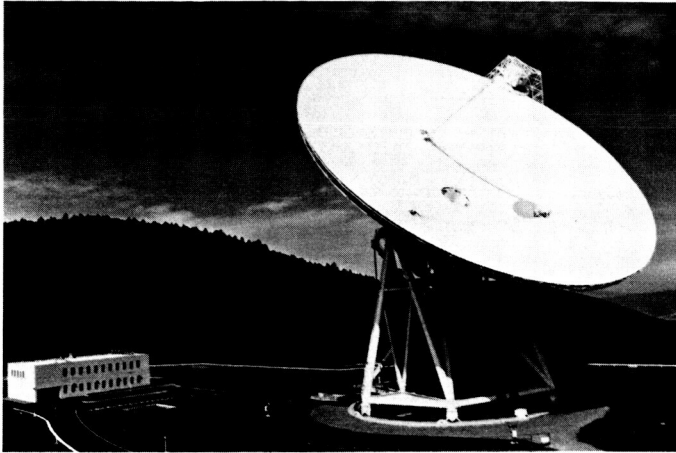
Fig. 5. Usuda 64-m antenna
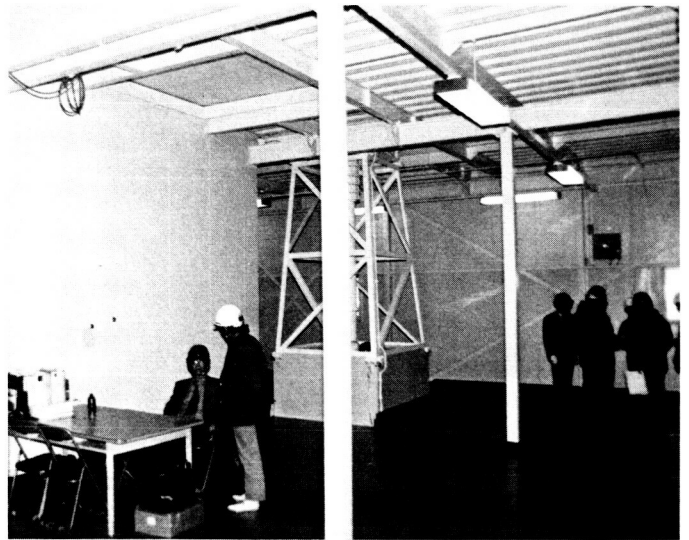
Fig. 7. Interior of Usuda alidade room

Fig. 6. Usuda 64-m antenna rear view

Fig. 8. 45-m radio astronomy antenna at Nobeyama, Japan

179
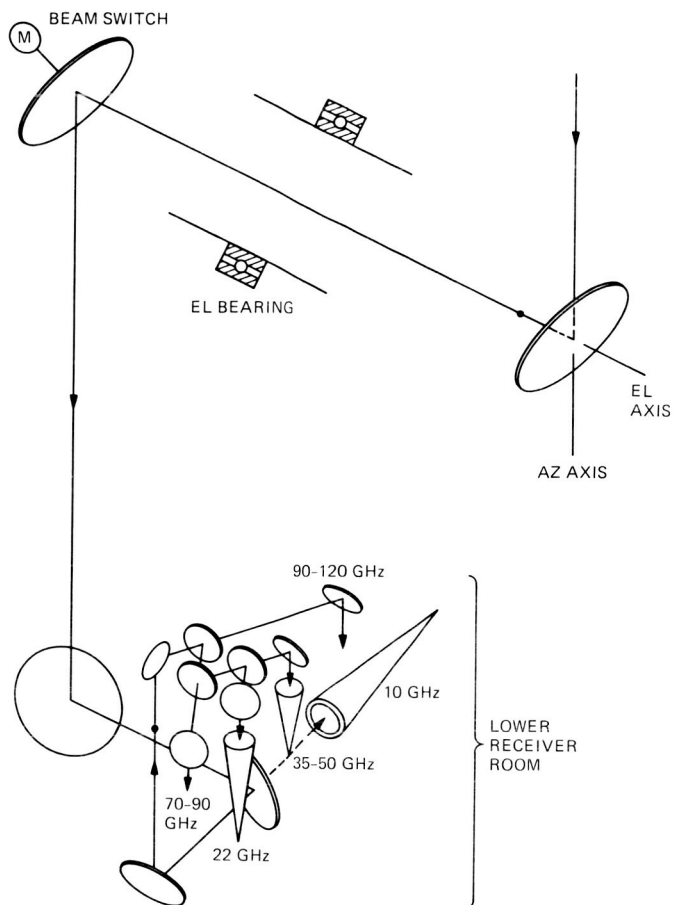
Fig. 9. Nobeyama 45-m antenna feed configuration

Fig. 11. Millimeter wave imaging sensing system

Fig. 10. Nobeyama 45-m antenna feed configuration and beam waveguide
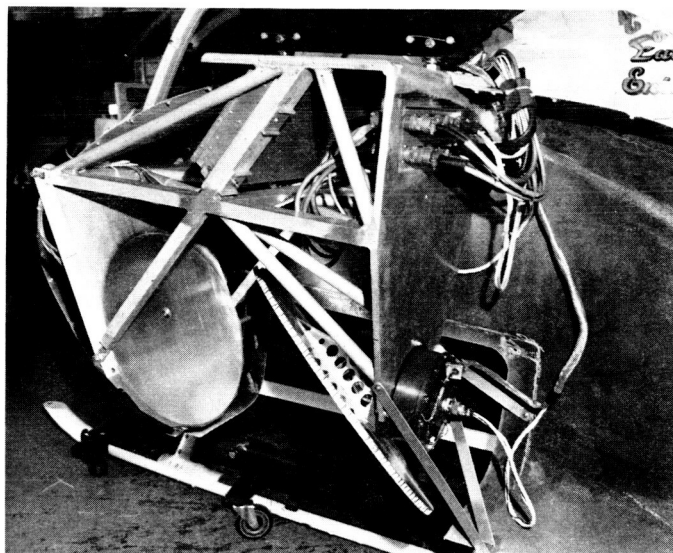
Fig. 12. Photograph of millimeter wave imaging sensing system antenna and scanning system
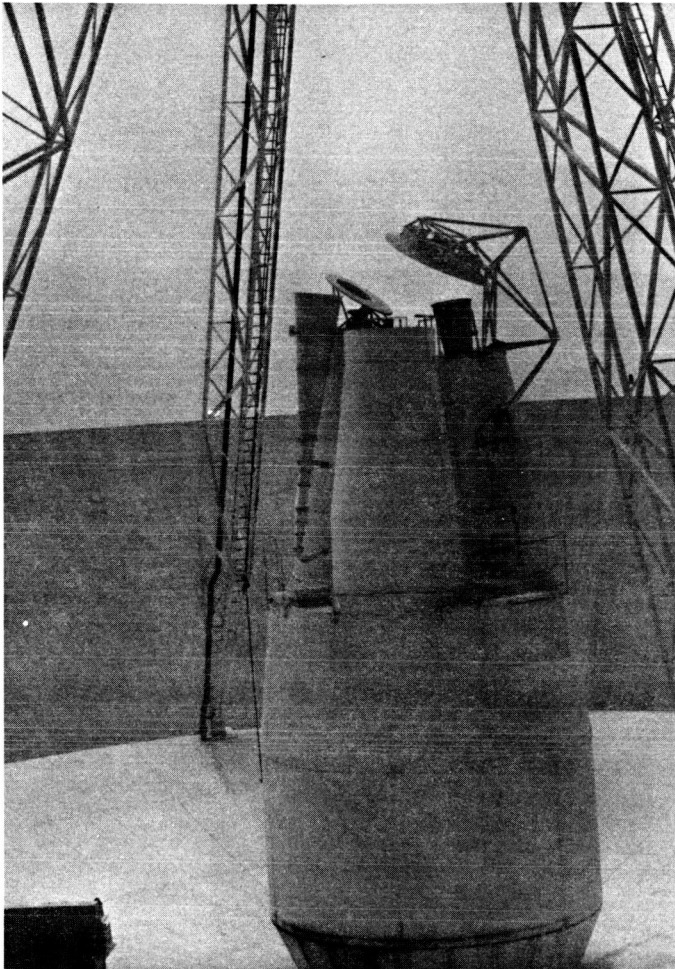
Fig. 13. DSS 14 64-m antenna feed system
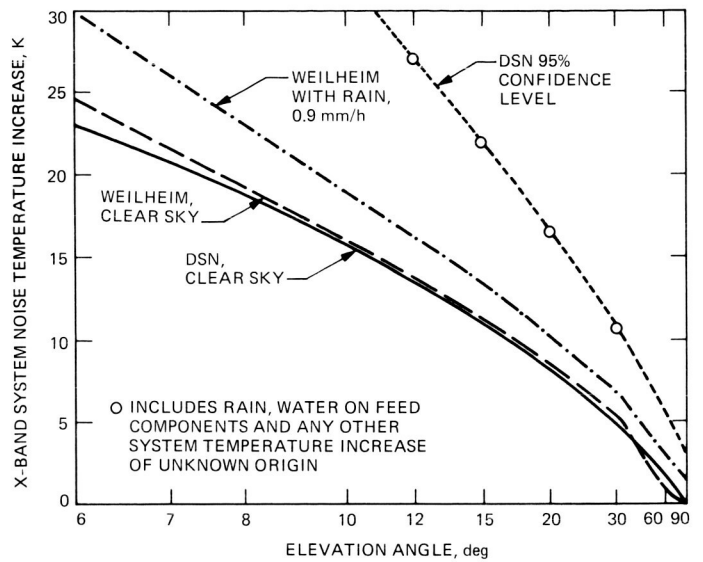
**Fig. 14. X-band system temperature increase versus
elevation angle**
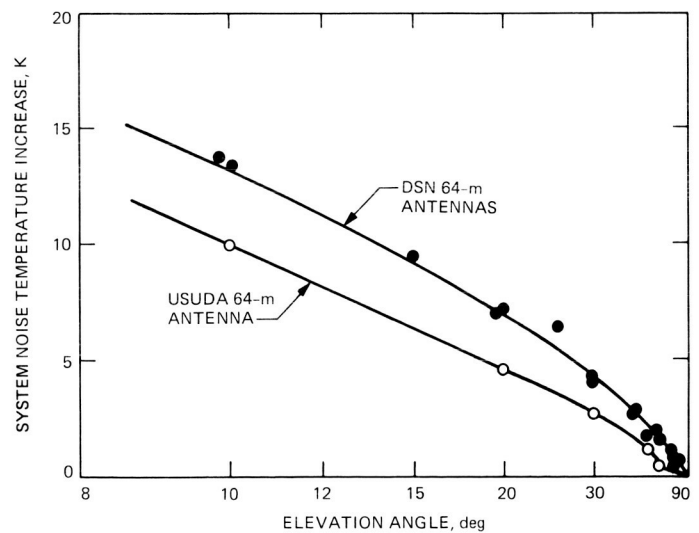
**Fig. 15. S-band system temperature versus antenna
elevation angle**
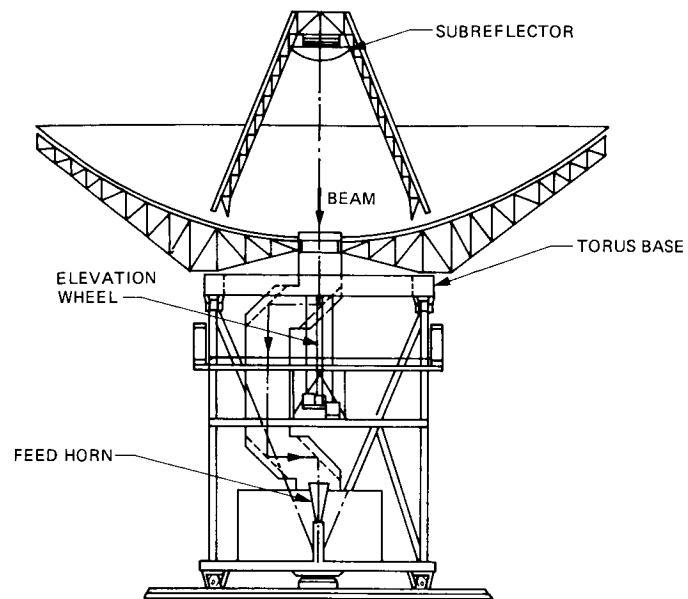
Fig. 16. Beam waveguide configuration for DSN 34-m HEF antenna
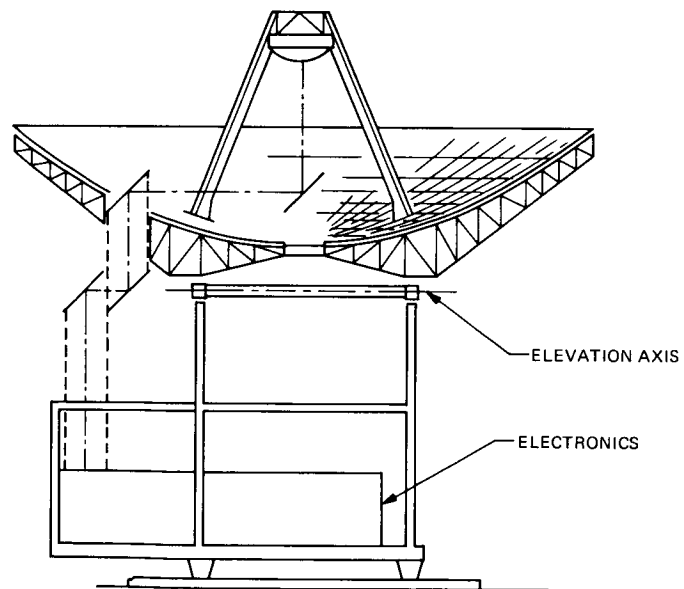
Fig. 17. Bypass-mode beam waveguide configuration for DSN 34-m HEF antenna

Structural Design Options for the New 34 Meter Beam Waveguide Antenna

S. Katow, J. J. Cucchissi, K. L. Chuang, R. Levy,
F. L. Lansing, F. W. Stoller and F. Menninger
Ground Antenna and Facilities Engineering Section

In addition to the successful network of 34-m High Efficiency antennas recently built by JPL, the Deep Space Network is embarking on the construction of a new 34-m high performance, research and development antenna with beam waveguide optics at the Venus site. The construction of this new antenna presents many engineering challenges in the areas of structural, mechanical, RF, and pointing system design. A set of functional and structural design requirements is outlined to guide analysts in the final configuration selection. Five design concepts are presented covering both the conventional center-fed beam optics as well as the nonconventional, by-pass beam configuration. The merits of each concept are discussed with an emphasis on obtaining a homologous design. The preliminary results of structural optimization efforts, currently in progress, are promising, indicating the feasibility of meeting, as a minimum, all X-band (8.4 GHz) requirements, with a goal towards meeting Ka-band (32 GHz) quality performance, at the present budget constraints.

I. Introduction

The planned construction of a new 34-meter diameter antenna with beam waveguide will enable the development of improved and flexible microwave optics, improved cryogenic equipment performance and maintenance, development of advanced transmitter and receiver and operations techniques, as well as the possibility of developing accurate antenna pointing hardware. It is also planned that the developed technology from this test bed antenna will be transferable to other antennas in the Network for increasing capabilities and performance improvement.

While this new high performance 34-meter antenna will closely resemble the present 34-m H. E. antennas in the Network, shown in Fig. 1, it presents many additional challenges in the areas of structural, mechanical, and microwave optics. For instance, the location of a beam waveguide near the center of the structure will necessitate the development of new elevation wheel-alidade configurations that accommodate the beam waveguide "tubes" without compromising structural performance.

In this article, the attention is focused on alternative structural design concepts and the merits of each with reference to the antenna functional requirements. Five design concepts emerged as candidates encompassing conventional, center-fed designs as well as nonconventional, by-pass designs. All concepts are presently undergoing extensive analysis and structure design optimization for performance evaluation prior to final decision making. Details of the concepts are given below.

II. Functional Requirements

From the structural-mechanical point of view the general functional requirements of the new antenna are tentatively scoped as follows:

(1) The antenna structure shall meet as a minimum all X-band (8.4 GHz) performance specifications under environmental loads; that is, it shall perform at least equal to the present 34-m High Efficiency antennas in the Network (Ref. 1). Subject to the funds available, the antenna design goal shall approach a Ka-band (32 GHz) quality performance.

(2) The antenna shall be an axisymmetric configuration with dual-shaped reflectors similar to the existing 34-m H. E. antennas.

(3) The antenna shall preferably be a center-fed beam waveguide (BWG) design with a built-in allowance for the addition of a by-pass BWG in the future.

(4) The antenna design shall be cost-effective in order to meet current Construction of Facility (COF) funding obligations.

In addition to the above general requirements, the specific design goals for the antenna structure are listed as follows:

(1) The gravity-loading path-length error (RMS) of the main reflector's backup structure shall be ⩽ 0.38 mm (0.015 inch). The Ka-band gravity RMS goal is 0.20 mm (0.008 inch).

(2) The wind loading path-length error (RMS) of the main reflector's backup structure shall be ⩽ 0.48 mm (0.019 inch) at 48 kph (30 mph) steady wind. The Ka-band wind RMS goal is 0.18 mm (0.007 inch) at 32 kph (20 mph).

(3) The wind-pointing error of the structure shall be ⩽ 13 milli degrees (mdeg) at 48 kph (30 mph) steady wind. The Ka-band wind-pointing error goal for the structure is 3 mdeg at 32 kph (20 mph) wind.

(4) The antenna surface panels shall be manufactured with a surface error tolerance (RMS) ⩽ 0.25 mm (0.010 inch). The Ka-band panel fabrication RMS goal is 0.13 mm (0.005 in.).

(5) The panel setting tolerance shall be at least equal to 0.25 mm (0.010 inch) with a goal of 0.13 mm (0.005 inch).

(6) The antenna structure shall survive steady wind loads up to 160 kph (100 mph) in the stow position.

(7) The antenna control system, servo drives, hardware and software shall, as a minimum, meet the X-band

(8.4 GHz) overall operational pointing system precision of 7 mdeg, with a goal of 2 mdeg for meeting the Ka-band (32 GHz) requirements. Both pointing requirements represent about 10% of the halfpower beamwidth at the corresponding frequency.

In addition to the above key requirements, many other requirements must be satisfied for antenna foundations, fire protection, safety, supporting facilities, monitor and control, and operation functions.

III. Candidate Concepts

The following five elevation wheel concepts were selected for investigation:

Concept A: Center-fed with "spokeless" elevation wheel with a torus at the main reflector base.

Concept B: Center-fed with homologous[1] double octagon elevation wheel base and a hollow elevation bearing.

Concept C: Center-fed with a split (two) elevation wheel.

Concept D: By-pass mode with an octagon elevation wheel similar to the existing 34-m H. E. octagon.

Concept E: By-pass mode with a homologous, double octagon-based elevation wheel.

Concepts D and E above are possible alternatives to current preferable center-fed concepts A thru C.

Each concept is described with the accompanying Figs. 2-8 as follows:

A. Concept A

In this center-fed, "spokeless" elevation wheel concept, the main reflector rests at four points connected to a rectangular cross-section torus as shown in Figs. 2 and 3, whose circumference is divided into 24 sectors. The torus is supported at the two elevation bearings. To ensure homology and axisymmetry, each of the four main reflector-torus interface connections lies on a radial rib each making 45 degrees with the elevation axis. The single elevation bull gear and counterweight lie in the antenna plane of symmetry, orthogonal to the elevation axis.

---

[1]The concept of a "homologous" antenna structure produces a structure that maintains, under varying gravity loadings, a perfect paraboloid surface (or a perfect shaped surface), although of different focal length, at all antenna tilts.

B. Concept B

In this center-fed hollow elevation bearing concept, shown in Figs. 4, 5 and 6, the main reflector is connected at eight points to the vertices of an octagon. Eight planar trusses join the octagonal base of an inverted pyramid which forms the elevation wheel. In this case, homology is achieved by locating the octagonal base in the plane of the elevation axis. The centerline of the BWG tube coincides with the elevation axis. In addition to housing the mirrors, the tube acts as a load-bearing element. The significant difference between this configuration and the 34-m H. E. antenna elevation is that the latter consists of only one plane at the octagon level while concept B possesses double octagon planes. Similar to the existing 34-m H. E. antenna elevation wheel, this concept has a single elevation bull gear with two eight-spoke inverted pyramids as well as counterweight located near vertex A of one inverted pyramid. The hollow bearing concept was also used in the 45-m antenna at Nobeyama, Japan (Ref. 2).

In order to minimize the distortion of the reflector's surface for gravity loading, the reflector combined with the basic elevation wheel structure must be axisymmetric about the central axis. Furthermore, the elevation wheel structure must be supported on two or more points on this axis and, in turn, the load must be transmitted by proper structural connections to the elevation bearings. This tipping structure combination must also be weight-balanced about the elevation axis (note that the tipping structure includes the main reflector, elevation wheel, subreflector, and its supports).

The cross-sectional areas of the structural members making up the tipping part of an antenna can then be designed by the JPL-IDEAS program to closely maintain an ideal surface. The computer program also allows focal length changes due to changes in the gravity loading components along the boresight axis (caused by rotation of the tipping structure).

The above structural design strategy, known as homology, has its best example in the Max Planck Institute's 100-meter Effelsberg antenna, near Bonn, Germany. This antenna uses a separate truss structure to carry the elevation gear and supports the reflector structure at two points on the reflector's truss structure. However, this concept of homology best addresses the gravity loading only but not wind loading. Application of the Effelsberg structure concept to the Deep Space Network (DSN) where low distortions under high wind velocities are necessary, results in excessive weight and thus a costly antenna.

The three high efficiency 34-m, X-band AZ/EL antennas now installed at the DSN stations have a partially homologous elevation wheel structure connecting the axisymmetric reflec-tor to the elevation wheel. A compromised truss configuration was arranged in order to avoid interference with two diagonal supporting bars of the alidade.

The solution to this interference problem, in addition to satisfying completely the homologous requirements, is now available in concept B. To satisfy the homology requirements, the octagon of the elevation wheel must be in the same plane containing the elevation axis. When the octagon plane is horizontal, the vertical reactions at the elevation bearings are equal to the load at apex A of the inverted pyramid which connects the corners of the octagon. This pyramid apex is one of two supporting points on the symmetric axis of the tipping assembly. Also necessary for the homology requirement is that the center of the cross bars be connected to the corners of the octagon.

The important structural members are the bars connecting the octagon to the elevation axis. These bars should only transmit axial forces and not carry bending moments. Usually, a bar with two flexures or easily bendable sections at its ends satisfies the above requirements. In addition, the main reflector is supported by vertical and sloped bars between the octagon and the eight points at the base of the reflector structure. This type of truss results in tangential support between the reflector and the octagon.

C. Concept C

In this center-fed split elevation wheel concept, two elevation drives are needed instead of one as shown in Fig. 7. The split wheel concept has been employed in other BWG antennas, such as in the 64-m antenna at Usuda, Japan (Ref. 3). The counterweight location will be divided into two equal parts, one at each elevation bull gear. One of the advantages of this concept is in its geometry, similar to the present 64-m/70-m antennas in the network. The technology developed after the construction of a beam waveguide antenna with this concept C may be directly transferred to the 64-m/70-m network in the future at no elevation wheel modification cost. On the other hand, because concept C has a completely different elevation wheel-drive arrangement, it can be viewed as costly in retrofitting the present 34-m H. E. antennas having a single elevation bull gear.

D. Concept D

In this bypass BWG concept, shown in Fig. 8, eight radial ribs of the main reflector backup structure are connected to the octagonal base of an inverted pyramid which forms the elevation wheel identical to the 34-m H. E. antenna of Fig. 1. The octagonal base is offset from the plane of the elevation axis by approximately 60.96 cm (2 ft) to provide clearance between the reflector backup trusses and the alidade struc-

ture. The eight vertices do not provide a homologous support for the backup structure. Four out of the eight points are directly connected to the elevation bearings by structural steel plates, thus making them more rigid than the remaining points. These plates are necessary to enhance structural rigidity and achieve the required wind pointing accuracy. This concept was originally investigated because it requires the fewest structural modifications to the present 34-m H. E. antennas.

E. Concept E

In this by-pass mode concept, homology will be achieved in the same manner as described in concept B (of Figs. 4, 5 and 6) except that the beam path will be outside the area encompassing the two elevation bearings as in concept D (of Fig. 8). The elevation bearings will be designed, in this case, to be the same as the 34-m H. E. antenna. This concept has, similar to concept D, the disadvantage of a costly incorporation of center-fed beam optics (unlike any one of the three concepts A, B or C) if added in the future.

IV. Structural Analysis Methodology

In an effort to determine the structural configuration that best meets the performance requirements, several candidate tipping structures were conceived. Finite-element models were developed for each of the above five concepts, and their member sizes (decision variables) are optimized under environmental (gravity and wind) loads. The objective was to find the lowest weight and minimum RMS structure that meets gravity pathlength error, wind pathlength error, and wind pointing accuracy requirements.

For each model, the main reflector backup structure starts with a geometry that is similar to the 34-m H. E. antenna; the important difference among the models is the elevation wheel configuration. The preferable elevation wheel configurations satisfy two basic criteria: (1) allowing unobstructed access for a BWG to the vertex region of the main reflector; and (2) providing equal stiffness (i.e., more homologous) supports for the main reflector backup trusses.

Structural analysis and optimization is performed using the JPL-IDEAS program (Refs. 4-7). Using a Lagrange multiplier-optimality criterion formulation, the algorithm finds optional objective function minimizations, such as the lowest structure weight under various environmental loads, while satisfying structure compliance constraints imposed on the antenna surface accuracy or boresight pointing performance. The two significant environmental loadings considered are gravity and wind. Gravity load consists of all structural and nonstructural dead load. Steady wind load, which results

from pressures applied to the antenna surface, is represented as resultant force vectors applied at the reflector backup structure nodes. The wind pressures, force and moment coefficients were derived from wind tunnel tests performed at Caltech[2] on representative scaled antenna models at several azimuth and elevation orientations.

Since the direction of the gravity loading vector relative to the structure varies over the elevation range of the reflector, the performance of the optimal design is rated at the antenna orientation(s) producing the worst gain loss. The gain loss is calculated from the Ruze equation, as proportional to the $(RMS)^2$ where RMS is the root mean square of the RF half-pathlength error of a paraboloid that best fits the deflected shape of the finite element grid points defining the antenna main reflector surface. The maximum gain loss occurs when the tipping structure is at either of the extreme elevation angles (0° or 90°). To mollify this effect, it is required that the RMS be zero (or minimum[3]) at a particular elevation angle, called the rigging angle, by forcing the reflector to be paraboloidal (or ideally "shaped") at this elevation angle. The rigging angle is determined from the IDEAS program by requiring equal RMS values at both zenith and horizon positions (Ref. 6), hence generating a single objective function to be minimized, instead of two.

Under wind loads, RF pathlength errors determined for the main reflector surface are adjusted to include error terms caused by shifts of the subreflector with respect to the position of the focal axis and nominal focal point of the best fit paraboloid. The equivalent adjustments are treated as additional, independent RMS pathlength errors. These offsets are not considered for gravity loading because the subreflector is automatically positioned to compensate for gravity deflections over the antenna elevation range.

Also, wind pointing error calculations include contributions from four components: (1) translation of best-fit paraboloid vertex, (2) rotation of best-fit paraboloid axes, (3) translation of the subreflector, and (4) rotation of the subreflector.

During optimization, a discrete set of values is chosen in the IDEAS program for sizing the design variables, which are

---

[2]R. B. Blaylock, "Aerodynamic Coefficients for a model of a paraboloidal reflector Directional Antenna proposed for a JPL Advanced Antenna System." Internal memorandum CP-6, Jet Propulsion Laboratory (internal document), Pasadena, Calif., May 1964.

[3]The 64-m antenna at Usuda, Japan realizes lower surface RMS at zenith and horizon looks with nonzero RMS at the rigging angle in between 0° and 90°.

the cross-sectional areas of truss elements. Size selection was made from a handbook of square structural steel tubes to achieve a realistic design

V. Results to Date

The results of the structural analyses to date are preliminary and are presented for each concept as follows.

A. Concept A

The torus concept described in Section III is the result of several configuration studies aimed at producing an efficient structure. Some of the options tried in connecting the main-reflector base to the torus are illustrated in Fig. 9.

Unlike the 34-m H. E. antenna elevation wheel, the torus concept does not have bars (or spokes) to transmit loads to the elevation bearings. Instead, the torus transmits loads directly to the elevation bearing via ring action. An initial option, used in an existing BWG antenna, has the torus connected to each of the 24 radial ribs. This nonhomologous support configuration did not perform as expected. Furthermore, it was adversely affected by increasing counterweight loads. The addition of interior stiffeners, or "bootstraps" as in Fig. 9, provided some limited performance improvement. Connecting the counterweight directly to the "bootstraps" rather than the torus yielded substantial improvement in gravity RMS relative to the structure weight. The design progress was hampered by the existing nonhomologous supports. To achieve better homology, the number of torus main-reflector backup connections was reduced from forty-eight (two per main rib) to only four. This approach has produced a design with one-third the gravity RMS, and two-thirds the structure weight of the preceding one. The pathlength and pointing errors under steady-state wind loads, on the other hand, exceed the performance criteria.

B. Concept B

The results made to date in analyzing concept E are mostly applicable to concept B. In order to use the homologous-type elevation wheel, with a few changes from existing 34-m H. E. antennas, the interference with an alidade member must be eliminated. This can be accomplished by moving the interfering diagonal members in the plane of the elevation axis which provide lateral support to the elevation bearings. Instead of these diagonal members joining the base of the alidade at the azimuth/radial bearing, they were moved to the rear center of the structure at either the base or at a higher plane. Rearranging some other diagonal members of the

alidade will be necessary to satisfy the function of the alidade.

A structural analysis was made using the JPL-IDEAS program and the existing 34-m H. E. antenna reflector computer model in combination with the homologous elevation wheel structure. Preliminary results indicated a gravity distortion (RMS) of the reflector of 0.10 mm (0.004 in.) for half pathlength at zenith or horizon position. When wind requirements are considered the design results in an increased weight and larger gravity distortions.

By increasing the height of the truss connecting the octagon to the reflector structure, adequate space is created to accommodate a large waveguide tube. Preliminary results showed that a 2.54 m (8.33 ft) round waveguide tube, as shown in Fig. 6, is possible.

Figure 6 shows also a square tube for the elevation axis shaft that does not violate homology conditions. Nodes A and B are the two points supporting the tipping assembly by a truss connection to the elevation bearings as required by homology. For supporting the tipping structure under the gravity loading component in the symmetric axis direction, node A is supported by rods AC and AD from the elevation axis bending-resistant shaft CD. This action requires that bars CE and CF only transmit axial forces and not bending.

For gravity loading support in the antisymmetirc direction, the two inverted cones from the octagon nodes to points A and B transfer the gravity loading forces to the elevation axis through bars CE and CF.

Increasing the distance between the octagon base and the elevation axis (as compared to the 34-m H. E. antenna reflector) to meet homology conditions has several unfavorable consequences that include the following:

(1) An increase in overturning moment on the alidade

(2) An elevation axis moment increase

(3) A yaw axis moment increase

A future change to a solid panel reflector surface would increase wind forces and moments. Improving the design for wind loading together with the cost impact are the subject of current investigations.

C. Concept C

The results of the split wheel concept are incomplete at the present time, but are expected to be close to either concepts A or B.

D. Concept D

The structural analysis of this concept has been completed, indicating that no substantial degradation of performance relative to the 34-m H. E. antennas will be expected.

VI. Summary

The construction of a new high performance research and development 34-m diameter antenna at JPL with beam waveguide optics presents many engineering challenges in the areas of structural, mechanical, RF, pointing and control system design. A set of functional and structure design requirements was outlined to guide analysts in the final configuration selection. Five design concepts were presented to include three conventional center-fed beam optics in addition to two nonconventional, bypass beam configurations. The merits of each concept were presented, with an emphasis on obtaining a homologous design. The status of the ongoing structural analysis and optimization effort for each concept was briefly discussed. Preliminary results are promising, indicating the feasibility of meeting, as a minimum, all X-band (8.4 GHz) requirements. Future work will include the selection of the final configuration to be built. The goal of satisfying a Ka-band (32 GHz) quality performance with the present budget constraints is also under investigation.

References

1. Pompa, M., "The new 34-m antenna," *Telecommunication and Data Acquisition Progress Report 42-85*, Jet Propulsion Laboratory, Pasadena, January–March 1986, pp. 127–138.

2. *The 45-m Telescope*, Nobeyama Radio Observatory Technical Report No. 6, Tokyo Astronomical Observatory, Tokyo, Japan, 1981.

3. Nomura, T. Hayashi, T. Nishimura, T. Hirosawa, H. and Ichikawa, M., *Usuda Deep Space Station with 64-m Diameter Antenna*, Institute of Space and Astronautical Science, Tokyo, Japan. 1985.

4. Levy, R., "Computer-Aided Design of Antenna Structures and Components," *Computers and Structures, 6*, 1976, pp. 419–428.

5. Levy, R., and Parzynski, W., "Optimality Criteria Solution Strategies in Multiple Constraint Design Optimization," *AIAA Journal, 20*, 5 1981, pp. 708–715.

6. Levy, R., "Antenna Rigging Angle Optimization Within Structural Member Size Design Optimization," in *Deep Space Network, Technical Report 32-1526, Vol. I*, Jet Propulsion Laboratory, Pasadena, CA pp. 81–87, Nov. 1970.

7. Utku, S., and Barondess, S. M., "Computation of Weighted Root-Mean-Square of Path Length Changes Caused by the Deformations and Imperfections of Rotational Paraboloidal Antenna," *Technical Memorandum 33-118*, Jet Propulsion Laboratory, Pasadena, CA, March 1963.
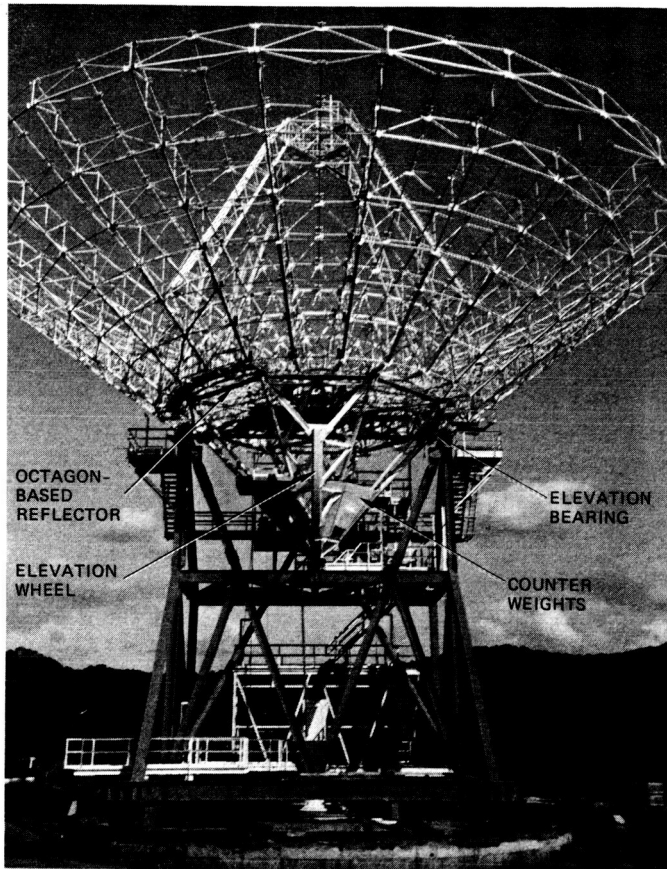
Fig. 1. The NASA/JPL 34-m H. E. antenna during construction
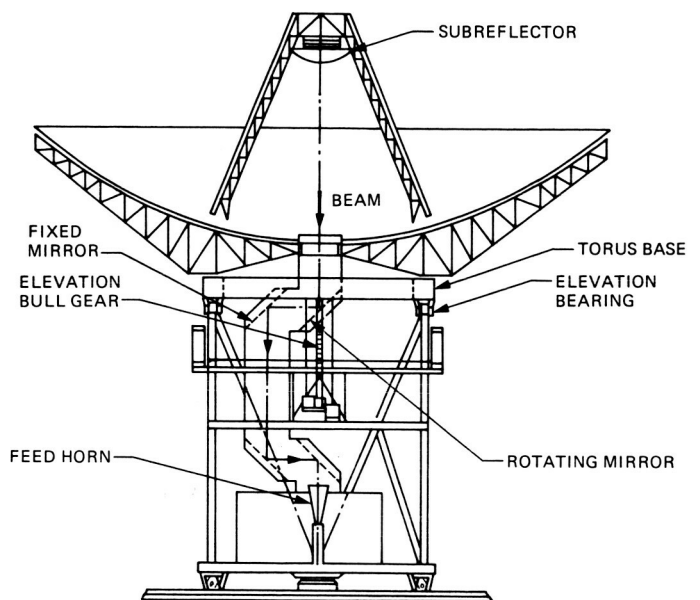
Fig. 2. A beam waveguide antenna layout with torus base,
Concept A
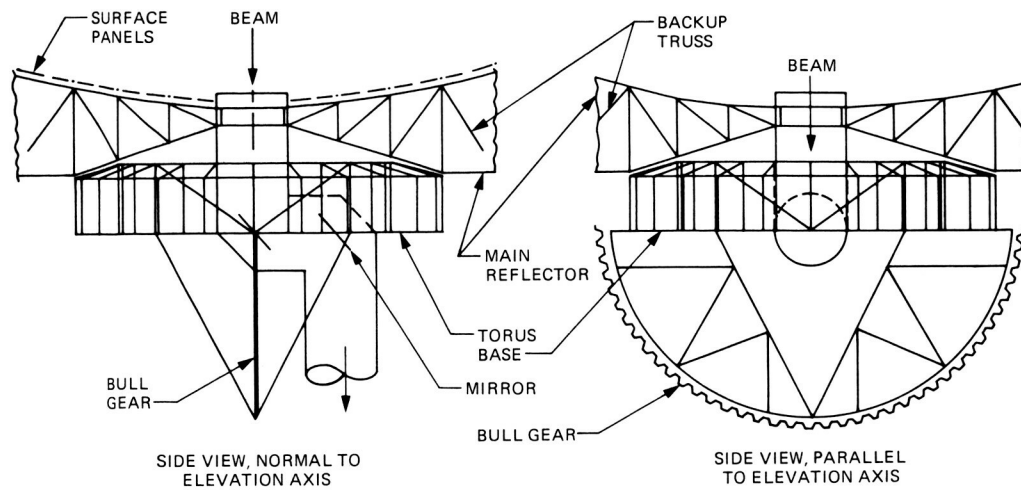
Fig. 3. Details of elevation wheel with torus base
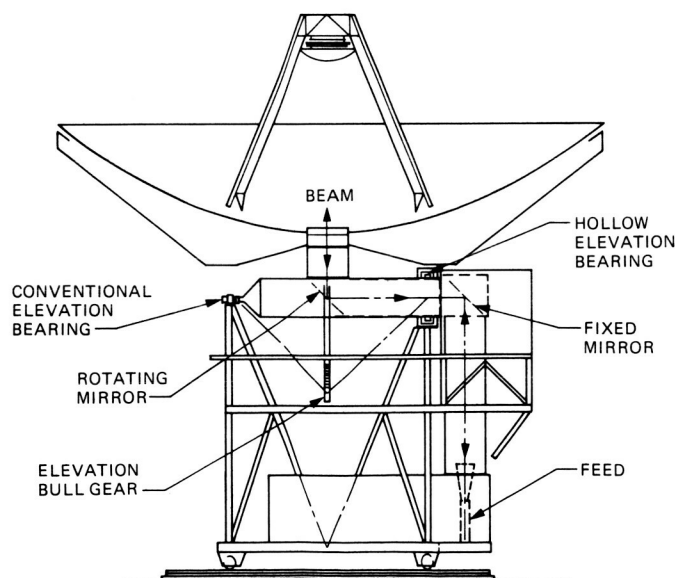
Fig. 4. A beam waveguide antenna layout with hollow elevation bearing tube, Concept B
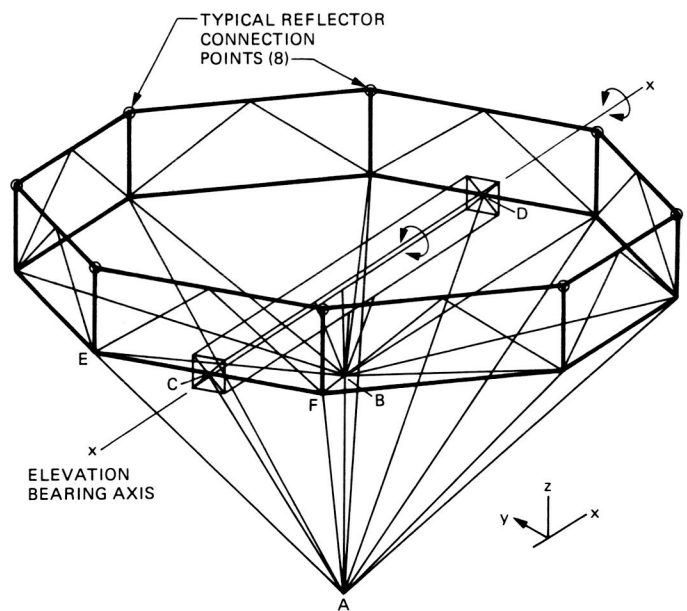
Fig. 5. Elevation wheel isometric view, Concept B

2.53 m (8.33 ft)
DIAMETER BEAM
WAVEGUIDE TUBE

C
E   F
B

A

SIDE VIEW, PARALLEL
TO THE ELEVATION AXIS

MAIN REFLECTOR BACKUP
SURFACE PANELS

x   D   x
C
ELEVATION
AXIS
B

HOLLOW BEAM
WAVEGUIDE TUBE

A

SIDE VIEW, NORMAL TO
THE ELEVATION AXIS
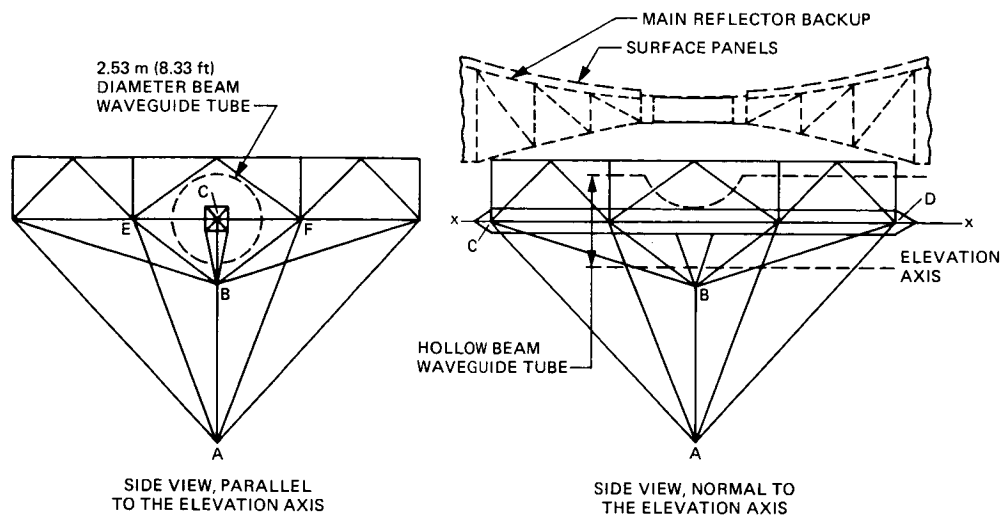
Fig. 6. Details of elevation wheel of Concept B

ELEVATION
BEARING

ELEVATION
DRIVE

REAR VIEW, HORIZON-LOOK

ELEVATION
AXIS
COUNTERWEIGHT

SIDE VIEW

TWIN
ELEVATION
BULL GEARS
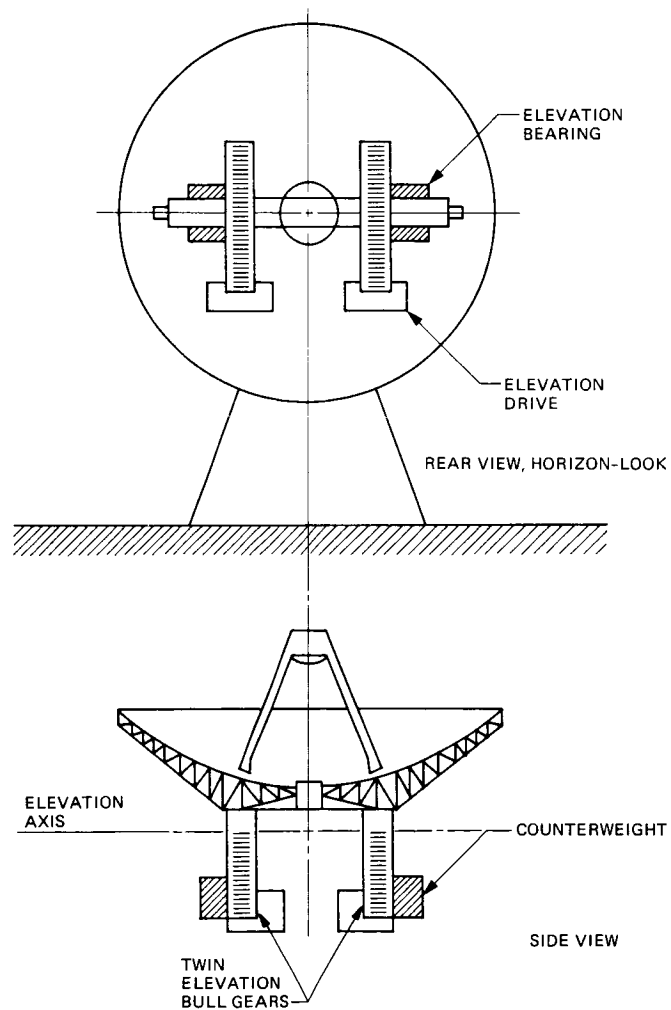
Fig. 7. A split elevation wheel concept, Concept C
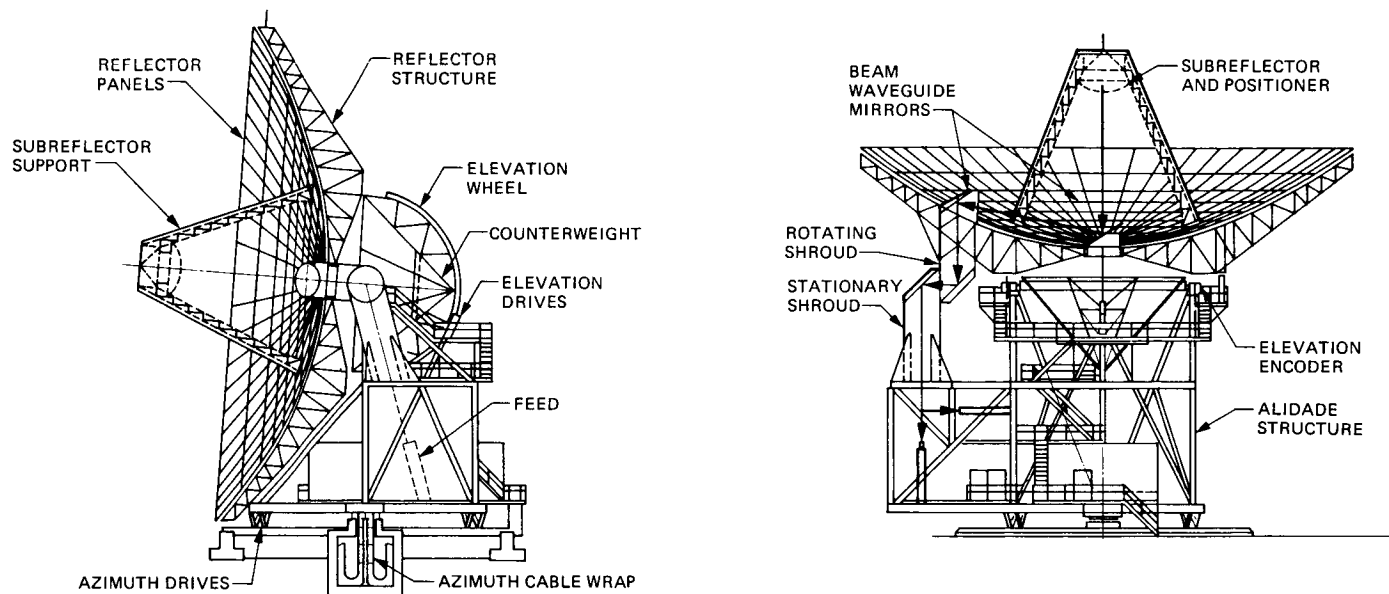
Fig. 8. A beam waveguide antenna with nonconventional, by-pass beam optics, Concept D
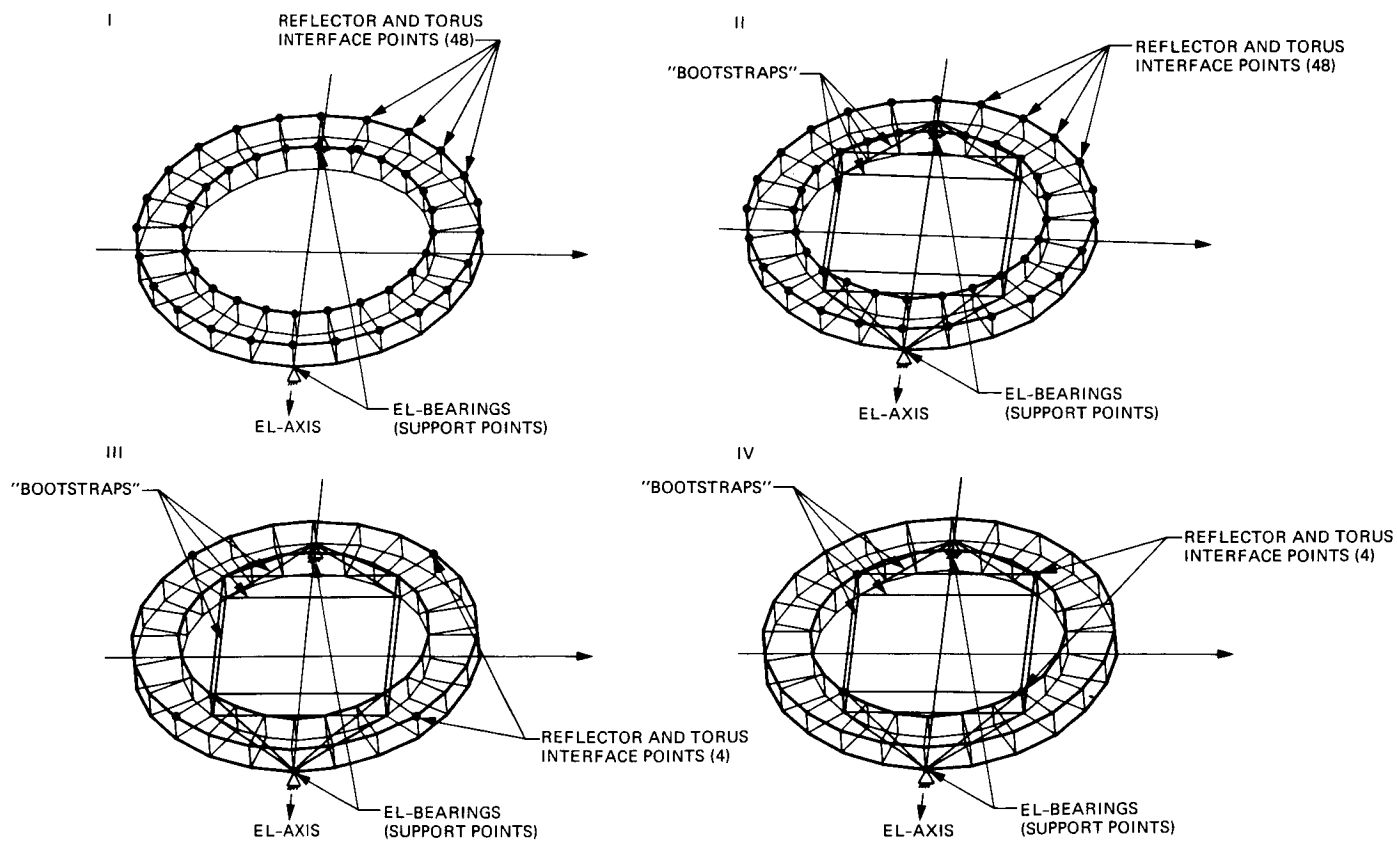
Fig. 9. Alternative main reflector-torus connectivity options

An Overview of the Mark IVA Monitor and Control System

J. G. Leflang

TDA Engineering

The Deep Space Network Monitor and Control System was completely changed during the conversion from the Mark III to the Mark IVA configuration. The new configuration employs shared data processing equipment between several co-located antennas, and incorporates much greater centralization of operations functions. The new configuration is described and its performance is compared to that of the Mark III era.

I. Introduction

In 1985, the Deep Space Network (DSN) was upgraded from the Mark III configuration to the Mark IVA configuration. The objectives of this upgrade were: (1) to replace aging equipment in the DSN, and (2) to change from stand-alone stations (each with one antenna) to a Signal Processing Center (SPC) fed by several co-located antennas.

To effect the upgrade, the entire Monitor and Control System (MON) was replaced. This was done in order to establish centralized control of all equipment in each of five facilities. These are: (1) The three Deep Space Communications Complexes (DSCCs) with their Signal Processing Centers, (2) the ground communications Central Communications Terminal (CCT), and (3) the Network Operations Control Center (NOCC).

Centralized control is considered to be essential to enable complicated operations activities and to control operational costs. The SPC configuration was selected because it requires less equipment at each DSCC. For example, if there are three stand-alone stations, *six* telemetry processors are required (one on-line and one for backup at each station), in order to meet the functional availability requirement. In the centralized SPC configuration, however, only *four* telemetry processors are required (one on-line for each station, and one which is a backup for all three stations).

The Mark IVA MON (see Fig. 1), with a few exceptions, provides the capabilities for operating subsystems at each facility from a central point. It acquires data to allow an operator to monitor the status and performance of the subsystem and allows centralized control of subsystem operation. The control function provides the capability for initialization, configuration, calibration, mode selection, and shutdown of DSN equipment. Some systems (such as Telemetry), however, are composed of subsystems which reside in more than one facility and, therefore, are not centrally controlled.

In addition to centralized facility control, the MON also provides operator consoles for coordination and monitoring of the Deep Space Network. Network monitoring includes generation of Network Performance Records (NPR) and displays for evaluating performance of the stations and delivery of received data. For network control, the MON distributes predictions, sequences of events (SOEs), schedules, and stan-

dards and limits (S&L) to the stations for use in facility operations. At the facility level, the MON distributes control directives and provides displays of responses, status, performance parameters, configuration data, logs, and failure and diagnostic data to operations and maintenance personnel.

Centralization of the monitor and control of subsystems at the DSCCs was accomplished by the utilization of minicomputers or microprocessor-based controllers within subsystems for interfacing with the DSCC Monitor and Control Subsystem (DMC).

The Mark IVA configuration of the MON includes new operator terminals along with new subsystem interfaces and displays. Real-time operator input/output (I/O) devices at the subsystems are specifically excluded for all new equipment implemented into the network.

At the DSCCs, a new Antenna Pointing Subsystem (APS) with an electrical interface has replaced the former APS with its papertape-driven operation and associated problems. Also, a minicomputer-based Test Support Assembly (TSA) has been added to generate telemetry simulation streams under central operator control. The TSA replaces the Simulation Conversion Assembly (SCA) formerly used.

II. Functional Allocations Within the Monitor and Control System at the Facility Level

Facility level MON functions are allocated to the DSCC, CCT, and NOCC.

Each subsystem controller within a facility interfaces with and responds to control of the associated facility controller. The subsystem provides the capability for self-monitoring and supplying the facility controller with status, performance, configuration, failure and diagnostic data. In turn, the facility controllers send facility configuration, status, and data flow information to the NOCC for presentation to the Network Operations Control Team.

III. Functional Allocations of the Monitor and Control System Within the Facilities

A. Deep Space Communication Complex

The monitor and control functions at the DSCC can be divided into three areas, (1) the DSCC Monitor and Control (DMC), (2) the DSCC subsystem processors and controllers, and (3) the Local Area Network (LAN). The functions of the DMC and the DSCC subsystem processors and controllers support the concept of the station's centralized monitor and control. The keyboards and display devices necessary to perform pre-pass readiness tests, scheduled tracking activities, post-pass procedures, and various tests to verify proper operation are located at a central control and display position at the DSCC. The configuration also permits centralized control of simultaneous tracking and data playback activities.

1. The DSCC Monitor and Control Subsystem (DMC). One of the functions of the DMC is to store the support data (predictions, schedule, SOE, and S&L) received from the NOCC so that operations can be conducted using these data. After assembling the DSCC equipment into links (strings of equipment required to support a spacecraft pass), the operator at each link console can initialize the assembled equipment and operate the link. The DMC operators (via input/output devices) monitor the performance and data accumulation and the operations of each subsystem at the DSCC.

Each DSCC generates an operations log containing inputs, responses, alarms and events in addition to generating and recording an Original Data Record (ODR) containing data products generated at the station.

DMC operations are divided into two separate areas as follows:

(a) *Complex Monitor and Control.* The Complex Monitor and Control (CMC) processor provides the operator interface for the assembling of DSCC resources into links for mission support. In addition, the CMC, (1) generates configuration and status data for facility and network monitoring, (2) receives support data from the NOCC, and (3) stores and distributes these data to the other subsystem controllers. Also, the CMC receives, processes, and displays event/alarm messages and maintains an operations log. The configuration of the DMC is such that the CMC is not in line with the data flow from the links. Hence the links (once established) are not affected by a CMC failure.

(b) *Link Monitor and Control.* The Link Monitor and Control (LMC) processors provide the operator interface for monitor and control of the equipment assigned as a link to support a particular tracking pass or operation. Like the CMC, the configuration of the LMC is such that if the LMC experiences a failure or shutdown, there is no impact to data flow in other systems.

The LMC accepts and processes status, performance, and configuration data from the other subsystem controllers and generates monitor data blocks for flight projects. The LMC also receives and displays event/alarm messages and display data from other subsystems. In addition, the LMC generates and maintains an operations log for operator, maintenance, and analyst use. This log includes operation directives, sub-

system responses, and event/alarm messages received from other subsystems.

2. DSCC Subsystem Processors and Controllers. The DSCC processors accept operator control inputs from the DMC, control and monitor their lower level assemblies, and generate responses, alarms, events, display and monitor data for use by the real-time operator. A self-test capability is incorporated into these processors.

3. Local Area Network. The Local Area Network (LAN) is the network through which the subsystems at the DSCC communicate. The DMC controls the routing map and monitors the activity of the LAN. The types of data monitored include the availability of, and the statistics for, each interface port.

B. Central Communications Terminal (CCT)

The Central Communications Terminal (CCT) Monitor and Control Subsystem (CCM) provides the man-machine interface for CCT facility operation through control of routing, processing, display of status and performance data, and generation of an operations log. The CCT also transmits monitor data to the NOCC for use by the Network Operations Control Team (NOCT).

In the same manner as at the DSCCs, the CCM sends control data to lower level assemblies and receives responses, status, and performance data from them for processing, display, and generation of data to be forwarded to the NOCT.

C. Network Operations Control Center

The NOCC Monitor and Control Subsystem (NMC) and the NOCC Support Subsystem (NSS) combine to provide control of the DSCCs. The NSS generates and distributes support data including predictions, DSN operations sequences, and schedule data for the DSCCs. The NMC provides the man-machine interface for control of other Real-Time Monitors (RTM) which monitor and report their system's configuration and performance to the NOCT. In addition, the NMC receives data from facility controllers and provides NOCT personnel with details of the performance, status, and configuration of the DSN. All systems data flow are monitored by NOCT personnel to ensure that mission support requirements are met.

The basic functions of the NOCC MON are further defined in the following areas:

1. The NOCC Monitor and Control Subsystem (NMC). The NMC provides the operator interface for monitoring and controlling of the equipment at the NOCC (facility control) and for monitoring the operation of the DSCC and the CCT (network control).

(a) *Operator Interface.* The operator interface provides the capability for the Network Operations Control Team (NOCT) to monitor and analyze the status, performance, and operation of the DSN without being in the data stream between the spacecraft and the project users. The NOCT operates in a Network Control Area which houses multiple consoles for use by the various NOCT personnel (Operations Chief, Track Controllers, etc.). Currently there are four consoles for Track Controllers. Each supports up to six projects simultaneously, with display selections at each console made independently of one another.

The NOCT coordinates the support of all scheduled activities and verifies that the DSN support meets the established and scheduled commitment. This coordination includes providing real-time interfaces with the flight projects and the DSCCs. The NOCT also assists with fault isolation and recovery, including reallocation of network resources.

(b) *NOCC Operation.* The Network Monitor Processor (NMP) is part of the NMC subsystem and is the interface hub for NOCC operations in that it receives and processes data from the DMC, CCT, other NOCC RTM subsystems. It also provides information, along with other RTMs, to the NOCT for real-time analysis of the status, performance and operation of the DSN. The processing functions of the NMP are performed in conjunction with other assemblies of the NMC which provide alphanumeric displays, graphic plots, printers, and the operator I/O devices.

The other NOCC RTMs (Command, Radio Science/VLBI, Telemetry, Tracking) monitor data received from subsystems at the DSCC and, in some cases, compare them to predictions provided by the NOCC Support Computer (NSC). The RTMs generate status, residuals, and display frames showing the configuration and performance of their associated system. These data are transmitted to the NMC for NOCT use. In addition, a System Performance Record (SPR), (an archival file of operations) is also generated by the RTMs.

(c) *Network Control.* The NMP provides for network control by the presentation of displays from the RTMs and by presentation of selected detail information from subsystems at the DSCCs.

IV. Hardware Configuration

A. Computers

1. Deep Space Communications Complex. The Modcomp II computers in use during the Mark III configuration have been retained for use by the ARA, CMD, DTM, and DTK

subsystems (see Fig. 1) during the Mark IVA era. New Mod-comp Classic 7845 computers have been implemented for use by the APS, DMC, DSP, and DTS subsystems. The DMC contains two computers (one prime and one backup) for the CMC and three computers for the LMC. Each of these computers uses an IEEE-488 interface for connection to its LAN-Network Interface Unit (NIU).

The DMC interfaces with the RCV, PPM, TXR, and FTS subsystems using (8080/8086) microprocessor-based controllers that have been added to these subsystems. These controllers are fitted with RS-232 interfaces for connection to their LAN-NIU.

The APS computer also interfaces with lower tier micro-processor-based controllers which drive the antennas.

2. Central Communications Terminal. Existing Mod-comp II computers have been retained for the Error Correction and Switching Assembly (ECS), the CCM, and the Network Communications Equipment (NCE). These computers control the routing of data among the Project Operations Control Centers (POCC), the NOCC, and the Mission Control and Computing Center (MCCC).

3. Network Operations Control Center. The Modcomp II computers used in the NOCC have been retained. Two VAX 11-780 computers have replaced the Sigma V computers previously used for the generation of support data. Four Star Switch Controllers (SSCs) provide the interconnection among the RTMs in the NOCC.

B. Mass Storage

1. Deep Space Communications Complex. With the exception of tape recorders used for generation of Original Data Records by telemetry, radio science/VLBI, and ARA processors, there is no provision for mass storage of data at the DSCC. The Modcomp II computers do have 2.5-Mb disk drives that are used for program loading and temporary storage of data. The Modcomp Classic computers have a 6.5-Mb cartridge/Winchester drive for program loading and temporary data storage. Program and temporary data storage (logs, predictions, etc.) in the CMC and APA is provided by 256-Mb disk drives attached to these units.

2. Central Communications Terminal. The hardware configuration in the CCT remains as it was in the Mark III configuration, with tape drives for the temporary storage of data records, and disk drives for program loading.

3. Network Operations Control Center. In the NOCC, 2.5-Mb cartridge disk drives have been retained for the Mod-

comp II computers. Each VAX 11-780 uses a 256-Mb disk drive for data storage.

C. Man-Machine Interface

1. Deep Space Communications Complex. While not centralized to the extent provided by the Mark IVA configuration, the Mark III DSN included a Data Systems Terminal (DST) for monitor and control of data system processors and a console for monitor and control of receiver equipment. At the DSCC, the DST and the Station Monitor and Control (SMAC) console were replaced by two consoles (prime and backup) and three LMC consoles (one per link). The CMC console has two keyboards (prime and backup), five 19-inch color graphic displays (three of which are normally active), and a voice subsystem interface.

The LMC console has one keyboard, three 19-inch color graphic displays, and a voice subsystem interface. Closed circuit TV monitors are mounted on top of these consoles for antenna surveillance.

At both the CMC and the LMC, the lower half of the console center screen is reserved for operator input, subsystem responses, event/progress advisory messages, and macro-procedure execution listings. The upper half is reserved for alarm messages. The remainder of the screen display areas are available for operator selected 1/4- or 1/2-screen displays that are generated by the DMC or other subsystems.

Local terminals are used for program loading at those computers that have disks (with the exception of the DMC). One of the remaining implementation items is the elimination of local consoles and terminals by establishing remote initialization of all computers and controllers.

Tape mounting and handling is still necessary for ODR recording. It is performed at recorders which are located in the computer area, not in the control room. Sometimes, the link operators load and label their own tapes, and sometimes a roving operator performs the task. The roving operator is assigned to those assemblies which require control and which have not been adapted for central control. Centralization of those controls is planned.

2. Central Communications Terminal (CCT). The CCT man-machine interface will remain as it was in the Mark III configuration. There are two consoles; one is operated by the Communications Chief (Comm Chief), and the other is operated by the Communications Technician (Comm Tech). Each of these consoles has keyboards, cathode-ray tubes (CRTs), telephones, and a voice subsystem interface. Both the Comm Chief and the Comm Tech have access to CCT status and configuration displays as well as NOCC-generated system displays.

3. Network Operations Control Center.

(a) *Network Operations Control Area.* Network Operations Control Area (NOCA) man-machine interface is similar to what it was in the Mark III configuration. There is a console for the Operations (OPS) Chief, four track-controller consoles, two analyst consoles, and a support console. Each of these consoles has CRTs, a display selector, a telephone, and a voice subsystem interface. There is a separate keyboard-CRT terminal that provides direct access to the RTMs in the Network Data Processing Area (NDPA).

Displays are generated by the RTMs and contain detailed performance and configuration data (e.g., TPA no., channel no., frame size, bit rate, and bit error rate). Graphic displays, such as the point plot of antenna pointing residuals, are also provided. It is a function of the operators to combine this information with schedules and operations sequences to ensure that the support provided is in accordance with the established commitment.

(b) *Network Data Processing Area.* The man-machine interface in the Network Data Processing Area (NDPA) is very similar to what it was in the Mark III configuration, although there has been some change in the hardware. Simple CRT terminals have replaced most of the Megadata Input/Output operator terminals, and dot matrix printers have replaced G.E. Terminets. The NDPA operators handle the loading and initialization of the RTMs. They also operate tape drives for generating data records.

D. Local Area Network

1. Deep Space Communications Complex. The LAN hardware configuration of the DSCC has been completely replaced by new technology. The previous 16-port SSCs have been replaced by an Ethernet CSMA-CD bus. (CSMA-CD is an abbreviation for Carrier Sense, Multiple Access, with Collision Detection.)

The LAN is composed of a coaxial cable bus which connects a network of transceivers, Network Interface Units (NIUs) and computers. The NIUs are microprocessor-based, and handle communication protocols with host computers and other NIUs. All of the Modcomp II, Classic computers, and microprocessor-based controllers are connected to the LAN. The computers and controllers exchange datagrams containing monitor, control, and system data. The system data include command, telemetry, tracking, radio science, VLBI, and telemetry simulation.

Because DSCC-10 is physically too large for a single Ethernet, the LAN at DSCC-10 has a bridge to another network at a remote station (DSS-12) which is 20 km away.

Each LAN is supported by a redundant Network Configuration Facility (NCF) that provides program storage (on disk) and is a server for downline loading of NIU software. Datagram routing is accomplished by the use of functional addressing, which allows substitution of equipment without the need for a change of destination coded by originators. The functional-to-physical address translation is accomplished in the NIU. The address translation table is loaded into all NIUs with one broadcast message.

In addition to the datagram service, the network provides virtual circuit capability. Virtual circuits are used to support printers at remote sites. Eventually, all printers and consoles will be connected using virtual circuits.

2. Central Communications Terminal. The Mark III, 16-port SSCs have been retained in the CCT. Three of these SSCs are connected to provide redundant operation.

3. Network Operations Control Center. The Mark III, 16-port SSCs have been retained in the NOCC. Four of these SSCs are connected to provide redundant operation.

V. Performance

A. Response Time

Because human operators are used at the DSCC, the response time of the computers and controllers reporting to the DMC computers is very important. The nominal response time has been established as one second. Many of the control directives are serviced within this period. There are some directives (such as moving the antenna to point), however, which cannot be serviced within this time. For those cases where processing operations require more than one second, the computer or controller immediately responds with a "processing" message and sends a completed "advisory" later.

B. Computers

The Modcomp II computers are using essentially the same software as was used during the Mark III configuration. They still perform fairly well, but it should be noted that some problems associated with real-time operations do occur. Some of these problems are occasional computer halts causing reloads and restarts. The computer memory (128-K bytes maximum) is very limited, and this makes program changes difficult.

The microprocessor-based controllers are generally successful. The 9600-baud interfaces are fast enough, there is plenty of memory available, and the reliability is high.

C. Staffing

One of the objectives of the Mark IVA implementation was to reduce the staffing level of the operations crews at the DSN facilities. This objective has been met with the centralization of the monitor and control functions at all the facilities. The central monitor and control concept, which provides fast and easy access in setting up and configuring links and associated systems, has been widely accepted by the real-time operators even though the work load on DSN operations has greatly increased. It should be noted that the decrease in operations staff required an increase in the systems knowledge of the individual operators since they became responsible for the operation of an entire link instead of only a portion of one.

D. Functional Availability

The functional availability requirements are unchanged: 99% for single antenna telemetry (96% in cruise mode) and 99.5% when more than one antenna is available for use.

At the time of the Mark IVA implementation, there was a step decrease from 99% to 96% in the telemetry system functional availability. There were many factors contributing to this decline, but a major problem appears to be a lack of familiarity with a totally new MON and LAN. Because the operations training program did not have an interactive simulator available, most of the training was accomplished on live spacecraft. The result of this was high operator stress and less than required performance.

Also, most of the subsystems did not exhibit adequate functional availability when they were first implemented, but subsequent testing and debugging of the software (in 1985) brought most of the subsystems back to the level of functional availability of the Mark III configuration. The control of antenna pointing was one of the most severe problems, but now appears to be adequately corrected.

The one-hour specification limit for routine pre-track calibrations remains unchanged for Mark IVA. After an initial learning period, where problems often led to excessive calibration times, the Mark IVA meets this requirement.

VI. Conclusion

While it is true that performance suffered from implementation problems and a relatively lengthy learning period of approximately one year, the DSN recovered from the transition to the SPC configuration and central control and greatly exceeded performance requirements in the support of the Voyager-Uranus encounter. The Mark IVA objectives were achieved without seriously degrading the high performance standards of the DSN, and thereby demonstrated that the concepts of centralized processing and control are viable means of improving resources utilization. The Mark IVA DSN, with less equipment per antenna and smaller staff, is providing a cruise mode telemetry functional availability of 98% which exceeds the requirement of 96% and compares well with the 99% experienced in the Mark III DSN configuration.

The above availability was experienced with software which has not significantly changed since November, 1985. Presently, a major software upgrade package is being delivered. It incorporates many anomaly corrections which will eliminate most of the problems which cause data loss due to antenna pointing errors, and will reduce the stress on operators who must now learn to work around many control anomalies. It is expected that experience gained next year will show functional availability at least as high as in the Mark III era. It also appears that the use of computers in a centralized control configuration will result in even higher performance.
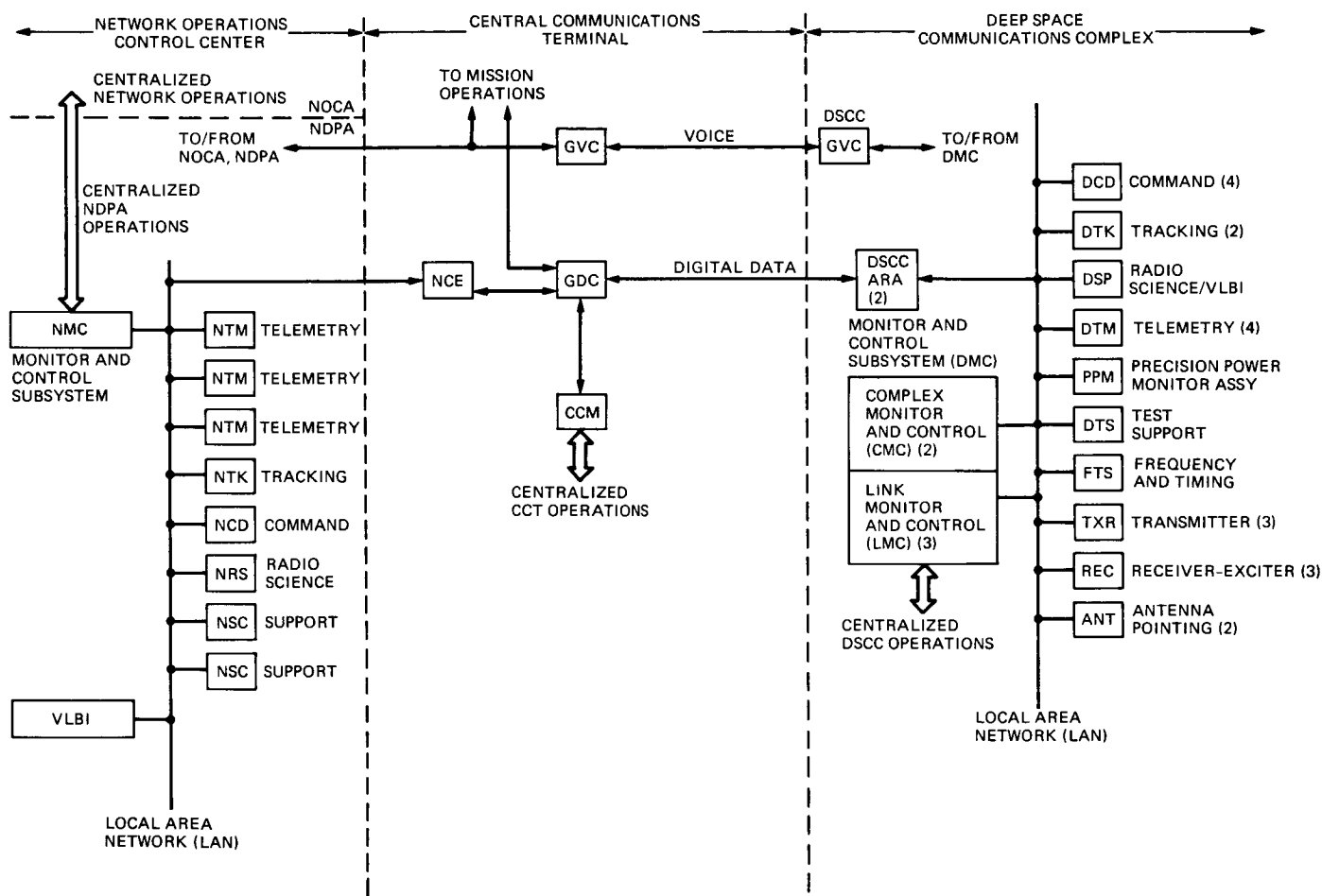
Fig. 1. DSN Monitor and Control System block diagram

Frequency Stability Review

C. A. Greenhall

Communications Systems Research Section

This tutorial article treats certain aspects of the description and measurement of oscillator stability. Topics covered are time and frequency deviations, Allan variance, the zero-crossing counter measurement technique, frequency drift removal, and the "three-cornered hat."

I. Deviations in Phase, Time, and Frequency

The purpose of this article is to define the Allan variance, relate it to various kinds of disturbances in oscillators and measurement systems, and to describe, from the point of view of an analysis and software person, the frequency stability measurement method used at the JPL Frequency Standards Laboratory. This is an expanded version of a talk given in Oct 1983 to DSN operations. It outlines the JPL frequency stability measurement methodology.

Let oscillators 1 and 2, both running at nominal frequency ν_0, have the outputs

$$\sin\left(2\pi\nu_0 t + \phi_1(t)\right), \quad \sin\left(2\pi\nu_0 t + \phi_2(t)\right)$$

where $\phi_1(t)$ and $\phi_2(t)$ are the phase deviations relative to frequency ν_0. Suppose that both oscillators are used for driving clocks. At true time t, the clock times t_1, t_2 are defined by

$$2\pi\nu_0 t_1 = 2\pi\nu_0 t + \phi_1(t) \tag{1}$$

$$2\pi\nu_0 t_2 = 2\pi\nu_0 t + \phi_2(t) \tag{2}$$

In other words, t_i is the time at which the output of a perfect oscillator, $\sin(2\pi\nu_0 t)$, would have the same total phase as oscillator i at time t.

The relative phase and time deviations of the pair of oscillators are defined by

$$\phi(t) = \phi_1(t) - \phi_2(t)$$

$$x(t) = t_1 - t_2$$

Subtracting (2) from (1) gives

$$x(t) = \frac{\phi(t)}{2\pi\nu_0}$$

The fractional frequency deviation is defined by

$$y(t) = \frac{1}{2\pi\nu_0} \frac{d\phi}{dt}$$

Then

$$y(t) = \frac{dx}{dt}$$

$$x(t) = \int y(t)\, dt$$

The average fractional frequency deviation over the time interval $(t-\tau, t)$ is defined by

$$\bar{y}(t,\tau) = \frac{1}{\tau}\int_{t-\tau}^{t} y(s)\, ds$$

$$= \frac{x(t) - x(t-\tau)}{\tau}$$

$$= \frac{\Delta_\tau x(t)}{\tau}$$

where Δ_τ is the backward first difference operator. We call τ the averaging time. When someone says that two oscillators differ in frequency by 5×10^{-14}, say, he is most likely quoting a value of $\bar{y}(t,\tau)$. To understand the quote, the value of τ must be known. Is it one second or one day?

Figure 1 shows plots of $x(t)$ and $\bar{y}(t,\tau)$ from a test of two hydrogen masers, where $\tau = 4094$ s and both x and $\bar{y}$ are sampled once every τ. The lower curve $\bar{y}$ is approximately the derivative of the upper curve x. Notice that $x = 0$ at both endpoints. This is an artifact of the presentation; the mean $\bar{y}$ has been subtracted off.

II. Allan Variance

The Allan variance was invented to solve the problem of characterizing the RMS frequency deviations of a pair of oscillators. Given time-deviation samples $x(0), x(\tau), \cdots, x(m\tau)$ over the time $T = m\tau$, let

$$\bar{y}_j = \bar{y}(j\tau, \tau)$$

the average frequency between $(j-1)\tau$ and $j\tau$. One could compute the averages

$$\langle x \rangle = \frac{1}{m+1}\sum_{j=0}^{m} x(j\tau)$$

$$\langle \bar{y} \rangle = \frac{1}{m}\sum_{j=1}^{m} \bar{y}_j$$

and the RMS deviations

$$x_{RMS} = \left[\frac{1}{m+1}\sum_{j=0}^{m}(x(j\tau) - \langle x\rangle)^2\right]^{1/2}$$

$$\bar{y}_{RMS} = \left[\frac{1}{m}\sum_{j=1}^{m}(\bar{y}_j - \langle\bar{y}\rangle)^2\right]^{1/2}$$

Often, however, these quantities tend to grow with T; moreover, they may fail to be reproducible from one T-interval to the next. (For example, if $\bar{y}$ is a random walk, then the mean and standard deviation of $\bar{y}_{RMS}^2$ are both proportional to T.)

For many types of oscillator phase deviations, the Allan variance, denoted by $\sigma_y^2(\tau)$, avoids these problems while still yielding a meaningful measure of frequency stability. To define theoretical Allan variance from the preceding setup, we must assume that $x(j\tau)$ is available for all non-negative integers j. Then, by definition,

$$\sigma_y^2(\tau) = \lim_{m\to\infty}\frac{1}{2(m-1)}\sum_{j=2}^{m}(\bar{y}_j - \bar{y}_{j-1})^2 \qquad (3)$$

if the limit exists. In other words, $\sigma_y(\tau)$ is the RMS average of the quantities $(\bar{y}_j - \bar{y}_{j-1})/\sqrt{2}$; thus $\sigma_y(\tau)$ measures the RMS change in $\bar{y}(t,\tau)$ when t changes by τ. Note that

$$\bar{y}_j - \bar{y}_{j-1} = \frac{1}{\tau}\left[\Delta_\tau x(j\tau) - \Delta_\tau x((j-1)\tau)\right]$$

$$= \frac{1}{\tau}\Delta_\tau^2 x(j\tau)$$

$$= \frac{1}{\tau}\left[x(j\tau) - 2x((j-1)\tau) + x((j-2)\tau)\right]$$

where Δ_τ^2 is the backward second difference operator. The quantity $\sigma_y^2(\tau)$ is called the Allan variance for the averaging time τ. It seems that $\sigma_y(\tau)$ (without the square) should be called Allan deviation. In this writeup, it will be done so for the sake of precision. In casual discussions, however, $\sigma_y(\tau)$ is called the Allan variance without confusing anyone.

The factor 2 (or $\sqrt{2}$) is there for historical reasons. On the other hand, if the two oscillators are judged to be of like quality, then the Allan deviation of the pair is sometimes divided by $\sqrt{2}$ to give the Allan deviation of the individual oscillators. This is a different $\sqrt{2}$. Moreover, this procedure assumes that the second τ-differences of phase of the two oscillators are

orthogonal random processes. If linear frequency drift is present, or if the two oscillators are subject to the same environmental fluctuations, then this assumption is unrealistic, at least for large τ.

The usual estimator of $\sigma_y^2(\tau)$, given $x(t)$ for $0 \leqslant t \leqslant T = m\tau$, is

$$S_y^2(\tau,m) = \frac{1}{2(m-1)} \sum_{j=2}^{m} (\bar{y}_j - \bar{y}_{j-1})^2$$

$$= \frac{1}{2\tau^2(m-1)} \sum_{j=2}^{m} \left[\Delta_\tau^2 x(j\tau)\right]^2 \qquad (4)$$

Table 1 gives a numerical example, a fragment from an actual test of two hydrogen masers. Elapsed time t is given in units of $\tau_0 = 256$ s, and $x(t)$ is given in units of 10^{-14} s. From these data, the following estimates can be calculated:

$$S_y(\tau_0, 8) = \frac{10^{-14}}{\sqrt{2}\,\tau_0} \left[\frac{1}{7}(87^2 + \cdots + 7^2)\right]^{1/2}$$

$$= 2.92 \times 10^{-15}$$

$$S_y(2\tau_0, 4) = \frac{10^{-14}}{\sqrt{2}\,2\tau_0} \left[\frac{1}{3}(125^2 + 56^2 + 37^2)\right]^{1/2}$$

$$= 1.13 \times 10^{-15}$$

$$S_y(3\tau_0, 2) = \frac{10^{-14}}{\sqrt{2}\,3\tau_0}\, 91$$

$$= 8.37 \times 10^{-16}$$

There are other estimators of $\sigma_y(\tau)$. The RMS average of the second differences of x can include values of $\Delta_\tau^2 x(t)$ for t not a multiple of τ. In the last example, for $\tau = 3\tau_0$, the values $\Delta_3^2 x(t)$, $t = 6, 7, 8$, can be used for computing the estimate

$$\frac{10^{-14}}{\sqrt{2}\,3\tau_0} \left[\frac{1}{3}(91^2 + 87^2 + 63^2)\right]^{1/2} = 7.48 \times 10^{-16}$$

Is this a better estimate than $S_y(3\tau_0, 2)$? If so, then in what sense is it better? This subject is controversial (Refs. 5, 6, and 8).[1]

---

[1] Also, D. Percival, letter to J. Barnes, Aug. 31, 1982.

Figure 2 shows $S_y(\tau, m)$ ("sigma") versus τ for the same test that yielded Fig. 1.

A. Naive Error Estimate

If, in the last sum of (4), the $\Delta_\tau^2 x(j\tau)$ were independent, zero-mean Gaussian random variables with the same variance, then $S_y^2(\tau, m)$ would be proportional to a χ^2 variable with $m - 1$ degrees of freedom. The assumptions of independence and zero mean are almost never realistic. Not knowing the true situation, however, one usually estimates the standard deviation of $S_y^2(\tau, m)$ by the naive formula $S_y^2(\tau, m)\epsilon$, where $\epsilon = \sqrt{2/(m-1)}$. Then

$$\left[S_y(\tau,m)\sqrt{1-\epsilon},\, S_y(\tau,m)\sqrt{1+\epsilon}\right] \qquad (5)$$

is presented as the roughest sort of "one-sigma" error bar for $S_y(\tau, m)$. If drift is removed (Section IV), then $\epsilon = \sqrt{2/(m-2)}$ is used. Figures 2 and 6 show the intervals (5) and also give $m - 1$, the "number of samples."

Don't use (5) unless $m \geqslant 4$; even then, don't take it seriously. Naturally, if you know the properties of the process $x(t)$ in advance, you can compute more accurate variance estimators and confidence intervals (Refs. 2, 3, 5, 6, 7, and 8).[1] But if you did know all this, you wouldn't be testing the oscillators in the first place. In any case, make sure that the user of the results knows the number of samples.

B. Deterministic Examples

The phase fluctuations of oscillators can often be modelled as a simple nonrandom function of time, plus a random component. Therefore, it is useful to know the effect of certain deterministic phase functions on Allan variance.

(i) *Constant phase and frequency offsets:*

$$x(t) = a_0 + a_1 t$$

Here, $\Delta_\tau^2 x(t) = 0$, so $\sigma_y^2(\tau) = 0$ for all τ. Such constant offsets have no effect on the Allan variance, which is non-zero only if the frequency difference of the two oscillators is changing with time.

(ii) *Linear frequency drift:*

$$x(t) = \frac{1}{2}ct^2 \qquad\qquad (\text{so } y(t) = ct)$$

In this case $\Delta_\tau^2 x(t) = c\tau^2$, and

$$\sigma_y(\tau) = \frac{1}{\sqrt{2}\tau}\,|c|\tau^2 = \frac{|c|\tau}{\sqrt{2}} \qquad (6)$$

A linear frequency drift causes the Allan deviation to be proportional to τ. Section IV shows one way of estimating the drift and removing it from the measurements.

(iii) *Higher powers of t.* If $n \geqslant 3$ and $x(t) = ct^n$, then

$$\Delta_t^2 x(t) = cn(n-1)\tau^2 t^{n-2} + \text{terms of lower degree}$$

Since this grows like a positive power of t, the Allan variance does not exist. (The limit in (3) is $+\infty$.)

(iv) *A single frequency spike.* This is the same as a step in phase or time. Suppose that $x(t)$ jumps by an amount X_0 during a measurement of duration T. Given $\tau = T/m$ there is an index k such that $\Delta_\tau x(k\tau) = X_0$ and the other $\Delta_\tau x(j\tau)$ are zero. Then

$$\Delta_\tau^2 x(kt) = X_0, \qquad \Delta_\tau^2 x((k+1)t) = -X_0$$

and the rest are zero. It follows from (4) that

$$S_y^2(\tau, m) = \frac{X_0^2}{\tau(T-\tau)}$$

Thus, for $\tau \ll T$, the estimated Allan variance is approximately proportional to $1/\tau$. A single frequency spike mimics white frequency noise (see below). This is appropriate, since a Poisson train of spikes *is* a form of white noise.

(v) *Periodic disturbances.* These can be caused by daily temperature variations, periodic weather fluctuations, problems with the measurement system, or even problems with the oscillators themselves. Their effect on Allan variance measurements can be bizarre. For example, let

$$x(t) = X_0 \cos(2\pi\nu t + \theta)$$

Some computation yields

$$\sigma_y(\tau) =$$

$$\frac{2}{\tau} |X_0| \sin^2(\pi\nu\tau) \cdot \begin{cases} \sqrt{2} |\cos\theta|, & \text{if } 2\nu\tau = \text{integer} \\ \\ 1, & \text{otherwise} \end{cases}$$

This is the theoretical Allan deviation (passing to the limit in (4)). For $\theta = 0$, this function looks like Fig. 3. If your Allan deviation plot has a lot of wiggles or

looks like a staircase, inspect your raw x or $\bar{y}$ data for periodic contamination.

C. The Classical Random Clock Noise Model

This is a combination of five "power-law" random processes, each of which contributes its signature to the Allan variance. The model is specified by the one-sided spectral density $S_y(f)$ of the fractional frequency $y(t)$:

$$S_y(f) = h_2 f^2 + h_1 f + h_0 + h_{-1} f^{-1} + h_{-2} f^{-2}$$

$$= \sum_{\alpha=-2}^{2} h_\alpha f^\alpha$$

Table 2 gives the Allan variances of the five components of the model (Ref. 1).

For example, if one sees $\sigma_y(\tau)$ proportional to $1/\sqrt{\tau}$ over two decades or more, one usually interprets this as the effect of white frequency noise. You had better eyeball the x or $\bar{y}$ data versus time to make sure that the $1/\sqrt{\tau}$ is not caused by a single monster phase jump (see item (iv) above). The inference of a model from $\sigma_y(\tau)$ is unreliable without some common-sense checks.

For white phase and flicker phase, the $\sigma_y^2(\tau)$ formulas are approximations that require $2\pi f_h \tau \gg 1$. Moreover, the white phase formula can be extended to the case of a stationary time deviation process $x(t)$ having an autocovariance function $R_x(\tau) = Ex(t)x(t+\tau)$. Let $\sigma_x^2 = R_x(0)$, the time variance. If there is a τ_0 such that $|R_x(\tau)| \ll R_x(0)$ for $\tau \geqslant \tau_0$, then

$$\sigma_y^2(\tau) \approx \frac{3\sigma_x^2}{\tau^2}, \quad \tau \geqslant \tau_0 \qquad (7)$$

III. Zero-Crossing Counter Technique

Figure 4 shows how the frequency and timing groups in the Communications Systems Research Section measure the time deviation $x(t)$ of two oscillators. The oscillators are offset in frequency by ν_b, which, in the present setup, is at most 1 Hz. Oscillator 2 runs at the higher frequency. If oscillators 1 and 2 have phases ϕ_1 and ϕ_2, then the beat-note signal is

$$\sin(2\pi\nu_b t - \phi(t))$$

where $\phi(t) = \phi_1(t) - \phi_2(t)$. The zero-crossing detector triggers the counter at the beat-note upcrossing times $t_0, t_1, t_2, \cdots$. The counter, running *continuously* at frequency ν_c (1 MHz in

the present setup), records discrete approximations t_0', t_1', t_2', $\cdots$ for further digital processing. The quantization errors $t_j' - t_j$ will be dealt with later.

Let $\tau_b = 1/\nu_b$, the nominal beat period. Then $t_j - t_{j-1} \approx \tau_b$. Note that τ_b is not precisely defined by the data. It could be $t_1 - t_0$, or the average of all the $t_j - t_{j-1}$ over the duration of the test. Changing τ_b also changes $\phi(t)$, but only by const $\cdot$ t, so that frequency stability is not changed.

The time deviation of the *beat note* is

$$x_b(t) = -\frac{\phi(t)}{2\pi\nu_b} = -\frac{\nu_0}{\nu_b} \frac{\phi(t)}{2\pi\nu_0} = -\frac{\nu_0}{\nu_b} x(t)$$

where $x(t)$ is the time deviation of the oscillator pair. (Note the sign reversal.) Since the factor ν_0/ν_b is large, typically 10^6 to 10^8, the fluctuations of the oscillators are magnified so that they can be measured. The Allan deviation of the oscillators is ν_b/ν_0 times the Allan deviation of the beat note.

The computation of the Allan variance of the beat note is not totally straightforward. By definition, the jth upcrossing t_j occurs when the total phase of the beat note is $2\pi j$, i.e.,

$$2\pi\nu_b t_j - \phi(t_j) = 2\pi j$$

At time t_j the time deviation of the beat note is

$$x_b(t_j) = -\frac{\phi(t_j)}{2\pi\nu_b} = j\tau_b - t_j, \quad j = 0, 1, 2, \cdots$$

(another sign reversal), and hence the time deviation of the oscillators is

$$x(t_j) = \frac{\nu_b t_j - j}{\nu_0}$$

Now suppose that the aim is to measure $\sigma_y(\tau)$ for $\tau = r\tau_b$, r a positive integer. This could be done if we had the sequence

$$x(t_0), \ x(t_0 + \tau), \ x(t_0 + 2\tau), \cdots \tag{8}$$

What we *do* have is the sequence

$$x(t_0), \ x(t_r), \ x(t_{2r}), \cdots \tag{9}$$

which is x at times that are not exactly spaced by τ. This is all right as long as the periods $t_{jr} - t_{(j-1)r}$ differ from τ by at most 1%, say. In other words, the fractional frequency stabil-

ity of the beat note must be no worse than 1%. The beat note itself must be a reasonably good clock, but not so good that its fluctuations are hidden by counter quantization. In most of our tests, this stability is 10^{-4} or better. There was one test, however, in which the beat periods $t_j - t_{j-1}$ fluctuated by a factor of two or more. In this situation, the Allan variance estimate given below is invalid. It is thought, however, that there might be a more sophisticated algorithm that would still be able to estimate Allan variance for $\tau \gg \tau_b$.

Since the second differences of j are zero, we have

$$\Delta_r^2 x(t_{jr}) = \frac{\nu_b}{\nu_0} \Delta_r^2 t_{jr} = \frac{\nu_b}{\nu_0} \left(t_{jr} - 2t_{(j-1)r} + t_{(j-2)r} \right)$$

Using this in (4) in place of $\Delta_\tau^2 x(j\tau)$ gives the Allan variance estimate

$$S_y^2(\tau, m) = \frac{\nu_b^2}{\nu_0^2} \frac{1}{2\tau^2} \frac{1}{m-1} \sum_{j=2}^{m} \left[\Delta_r^2 t_{jr} \right]^2$$

where $\tau = r\tau_0$ and the duration of the test is $m\tau$. Look again at the numerical example in Section II. The $x(t)$ column of Table 1 actually gives t_{jr}' in microseconds ($r = 256, j = 0$ to 8, $\tau_b = 1$ s) with the gross linear part $jr\tau_b$ already subtracted off. Since $\nu_b = 1$ Hz, $\nu_0 = 10^8$ Hz, the scale factor is 10^{-6} $\nu_b/\nu_0 = 10^{-14}$.

Quantization Error. Let the counter frequency be ν_c. (Ours is 1 MHz.) My model for the recorded time t_j' of the jth upcrossing t_j is

$$t_j' = t_j + q_j$$

where t_j' is a multiple of $1/\nu_c$, and q_j, the quantization error, satisfies $0 \leqslant q_j < 1/\nu_c$. As an approximation, assume that the processes t_j and q_j are uncorrelated, and that q_j is a process of independent, uniform random variables. This approximation also yields Sheppard's second-moment correction in statistics. Then the q_j alone contribute an Allan variance

$$\sigma_q^2(\tau) = \left(\frac{\nu_b}{\nu_0} \frac{1}{2\nu_c \tau} \right)^2$$

(This follows from (7).) For example, if $\nu_0 = 5$ MHz, $\nu_b = 1$ Hz, $\nu_c = 1$ MHz, then $\sigma_q(\tau) = 10^{-13}/\tau$. One should *subtract* $\sigma_q^2(\tau)$ from the measured $\sigma_y^2(\tau)$ (from t_j') to get the true $\sigma_y^2(\tau)$ (from t_j). In other words, $\sigma_q^2(\tau)$ contributes to the measurement-system noise floor.

IV. Drift Removal

During a long test of a pair of hydrogen masers, it is often apparent that the frequency deviation $y(t)$ appears to be dominated by a linear component, which also causes $\sigma_y(\tau)$ to increase like τ for large τ, thus masking the effect of the random fluctuations. See Fig. 5 and the upper curve of Fig. 6. The model is

$$x(t) = a_0 + a_1 t + \frac{1}{2} c t^2 + x_0(t)$$

or

$$y(t) = a_1 + ct + \frac{dx_0}{dt}$$

where c is the frequency drift rate ($\Delta f/f$ per second) and $x_0(t)$ is a mean-zero process. The constants a_0, a_1 are irrelevant for a frequency stability study. Let $\sigma_y^2(\tau)$, $\sigma_{y0}^2(\tau)$ be the Allan variances of $x(t)$, $x_0(t)$. We would like to estimate c and to use the estimate $\hat{c}$ to extract an estimate of $\sigma_{y0}^2(\tau)$, the Allan variance of the underlying fluctuations.

A. Estimating $\hat{c}$ From $x(t)$, $0 \le t \le T$

Assume a sample time of τ_0. The $\hat{c}$ estimators given below are all unbiased.

Method 0. Least-squares quadratic fit to $x(t)$:

$$\hat{x}(t) = \hat{a}_0 + \hat{a}_1 t + \frac{1}{2} \hat{c}_0 t^2$$

where $\hat{a}_0$, $\hat{a}_1$, $\hat{c}_0$ are chosen to minimize $\Sigma [\hat{x}(t) - x(t)]^2$. This method is optimal (minimal variance) in the presence of white phase noise.

Method 1. Least-squares linear fit to $\bar{y}(t, \tau_0)$:

$$\hat{y}(t) = \hat{a}_1 + \hat{c}_1 t$$

where $\hat{a}_1$, $\hat{c}_1$ are chosen to minimize $\Sigma [\hat{y}(t) - \bar{y}(t, \tau_0)]^2$. This method is optimal in the presence of white frequency noise.

Method 2. Mixed second difference of phase. This is the one that has been used here for tests of frequency standards (Refs. 4 and 7). No one else is thought to have used it. So far, no complaints have been received. Let

$$\hat{c}_2 = \frac{\bar{y}(T, \tau_c) - \bar{y}(\tau_c, \tau_c)}{T - \tau_c}$$

In other words, drift rate equals average frequency at the end of the run, minus average frequency at the beginning of the run, divided by the time span between the midpoints of the averaging intervals $[0, \tau_c]$ and $[T - \tau_c, T]$. Take

$$\tau_c = \frac{T}{6.29}$$

a value that minimizes the variance of $\hat{c}_2$ in the presence of flicker-frequency noise. This value is not critical. In terms of $x(t)$,

$$\hat{c}_2 = \frac{x(T) - x(T - \tau_c) - x(\tau_c) + x(0)}{\tau_c (T - \tau_c)} \tag{10}$$

so that only four values of $x(t)$ are needed.

This estimator, although suboptimal for all the classical noises, performs well in the presence of white, flicker, and random walk frequency noises. If $x_0(t)$ is random walk frequency, then the standard deviation of $\hat{c}_2$ is

$$\sigma(\hat{c}_2) = 4.6 \frac{\sigma_{y0}(\tau_c)}{T} \tag{11}$$

and this formula is pessimistic for white and flicker frequency. Note that you have to remove drift in order to estimate $\sigma_{y0}(\tau_c)$; see (13) below.

B. Removing Drift

Suppose that some unbiased estimate $\hat{c}$ is obtained, perhaps one of the foregoing, perhaps some other. An estimate of the residual $x_0(t)$ is

$$\hat{x}_0(t) = x(t) - \frac{1}{2} \hat{c} t (t - T)$$

See Fig. 7 for the effect of this ($\hat{c} = \hat{c}_2$) on the data of Fig. 5. Then

$$\Delta_\tau^2 \hat{x}_0(t) = \Delta_\tau^2 x(t) - \hat{c} \tau^2$$

is used in the estimate

$$S_{y0}^2(\tau, m) = \frac{1}{2\tau^2} \frac{1}{m-1} \sum_{j=2}^{m} \left[\Delta_\tau^2 \hat{x}_0(j\tau) \right]^2$$

for $\sigma_{y0}^2(\tau)$. Expanding the square gives

$$S_{y0}^2(\tau, m) = S_y^2(\tau, m) - \frac{\tau^2}{2} c_\tau^2 + \frac{\tau^2}{2} (c_\tau - \hat{c})^2 \tag{12}$$

where

$$c_\tau = \frac{x(T) - x(T-\tau) - x(\tau) + x(0)}{\tau(T-\tau)}$$

and $T = m\tau$. If $\hat{c} = \hat{c}_2$, then

$$S_{y0}^2(\tau_c,6) = S_y^2(\tau_c,6) - \frac{\tau_c^2}{2}\hat{c}_2^2 \qquad (13)$$

is used for estimating $\sigma_{y0}^2(\tau_c)$, needed in (11). The lower curve in Fig. 6 is $S_{y0}(\tau,m)$, the upper curve is $S_y(\tau,m)$, and the straight line is (6) with $c = \hat{c}_2$.

CAUTION

If $\hat{c}$ is based on the test data $x(t)$ $(0 \leqslant t \leqslant T)$, $T = m\tau$, m is small ($\leqslant 5$, say), and random walk frequency noise dominates $x_0(t)$, then $S_{y0}^2(\tau,m)$ severely underestimates $\sigma_{y0}^2(\tau)$ on the average, and has a larger variance than $S_y^2(\tau,m)$. If $\hat{c} = \hat{c}_2$, the bias of S_{y0}^2 is given by Ref. 7.

m	10	7	5	4	3	2
Bias	-16%	-24%	-34%	-43%	-59%	-89%

The error bars in Fig. 6 are the naive ones, and do not reflect these biases. The lesson here is to use a long-term estimate of c if possible. For example, cavity retuning of a hydrogen maser gives an independent measurement of frequency drift.

V. Three-Cornered Hat

The previous material has been about the relative stability of a pair of oscillators. If one has a triplet of oscillators, one may be able to estimate the stability of each one individually by using the three pair-comparisons. The setup, as shown in Fig. 8, has oscillators A, B, C, and pair-channels 1, 2, 3. We measure the pair Allan variances $\sigma_1^2(\tau)$, $\sigma_2^2(\tau)$, $\sigma_3^2(\tau)$ and would like to compute $\sigma_A^2(\tau)$, $\sigma_B^2(\tau)$, $\sigma_C^2(\tau)$ for the individual oscillators.

Assumptions

(1) The phase fluctuations of the three oscillators are independent.

(2) Relative linear frequency drift is negligible or removed. In other words, the second differences of x_A, x_B, x_C have mean zero.

If the assumptions are satisfied, then, for a fixed τ, the processes $\Delta_\tau^2 x_A$, $\Delta_\tau^2 x_B$, $\Delta_\tau^2 x_C$ are orthogonal. Therefore,

$$\sigma_1^2 = \sigma_B^2 + \sigma_C^2$$

$$\sigma_2^2 = \sigma_C^2 + \sigma_A^2 \qquad (14)$$

$$\sigma_3^2 = \sigma_A^2 + \sigma_B^2$$

and so

$$\sigma_A^2 = \frac{1}{2}\left(\sigma_2^2 + \sigma_3^2 - \sigma_1^2\right)$$

$$\sigma_B^2 = \frac{1}{2}\left(\sigma_3^2 + \sigma_1^2 - \sigma_2^2\right) \qquad (15)$$

$$\sigma_C^2 = \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2 - \sigma_3^2\right)$$

Figure 8 shows the geometric interpretation.

Difficulties

(1) Since frequency standards are sensitive to changes in the environment (temperature, pressure, humidity, magnetic field), it seems that frequency standards in the same room must violate assumption 1 for long-term fluctuations, say, for 1000 s or more.

(2) The pair-channel $\sigma^2(\tau)$ values are not available, just estimates such as $S^2(\tau,m)$. When estimates are substituted for the σ^2 values in the right side of (15), it often happens that some oscillator σ^2 comes out negative. One interpretation of this is that a confidence interval for that σ goes all the way down to zero. For example, if $S_1^2 > S_2^2 + S_3^2$, then the σ_A^2 estimate is negative. (If a triangle has sides S_1, S_2, S_3, then the angle between S_2 and S_3 is obtuse.) If the stronger inequality $S_1 > S_2 + S_3$ holds, then something is wrong with the measurements. (There is no triangle with sides S_1, S_2, S_3.)

References

1. Rutman, J., "Characterization of phase and frequency instabilities in precision frequency sources: fifteen years of progress," *Proc. IEEE*, Vol. 66, pp. 1048–1075, 1978.

2. Lesage, P., and Audoin, C., "Estimation of the two-sample variance with limited number of data," *Proc. 31st Ann. Symp. on Freq. Control*, Document No. ADA088221, National Technical Information Service, Springfield, VA, pp. 311–318, 1977.

3. Yoshimura, K., "Characterization of frequency stability: uncertainty due to the autocorrelation of the frequency fluctuations," *IEEE Trans. Instrum. Meas.*, Vol. IM-27, pp. 1–7, 1978.

4. Greenhall, C., "Removal of drift from frequency stability measurements," *TDA Progress Report 42-65*, pp. 127–132, Jet Propulsion Laboratory, Pasadena, California, 1981.

5. Greenhall, C., "A structure function representation theorem with applications to frequency stability estimation," *IEEE Trans. Instrum. Meas.*, Vol. IM-32, pp. 364–370, 1983.

6. Percival, D., *The Statistics of Long Memory Processes*, Ph.D. Thesis, U. of Washington, 1983.

7. Greenhall, C., "The fundamental structure function of oscillator noise models," *Proc. 14th Ann. PTTI Meeting* (NASA Conference Publication 2265), pp. 281–294, 1982.

8. Howe, D., et al., "Properties of signal sources and measurement methods," *Proc. 35th Ann. Freq. Control. Symp.*, Electronic Industries Association, Washington, DC, pp. A1–A47, 1981.

Table 1. Data from a stability test of two hydrogen masers

t	$x(t)$	$\Delta_1 x(t)$	$\Delta_1^2 x(t)$	$\Delta_2 x(t)$	$\Delta_2^2 x(t)$	$\Delta_3 x(t)$	$\Delta_3^2 x(t)$
0	0						
1	658	658					
2	1229	571	−87	1229			
3	1701	472	−99			1701	
4	2333	632	160	1104	−125	1675	
5	2991	658	26			1762	
6	3493	502	−156	1160	56	1792	91
7	4095	602	100			1762	87
8	4690	595	−7	1197	37	1699	−63

Table 2. Spectral density and Allan variance of the components of the classical clock noise model

α	Name	$S_y(f)$	$\sigma_y^2(\tau)$
2	White phase	$h_2 f^2,\ f < f_h$	$h_2 f_h \dfrac{3}{4\pi^2 \tau^2}$
1	Flicker phase	$h_1 f,\ f < f_h$	$h_1 \dfrac{3}{4\pi^2 \tau^2} \ln(8.88 f_h \tau)$
0	White frequency	$h_0,\ f_h = +\infty$	$h_0 \dfrac{1}{2\tau}$
−1	Flicker frequency	$h_{-1} f^{-1},\ f_h = +\infty$	$h_{-1} \ln 4$
−2	Random walk frequency	$h_{-2} f^{-2},\ f_h = +\infty$	$h_{-2} \dfrac{2\pi^2 \tau}{3}$
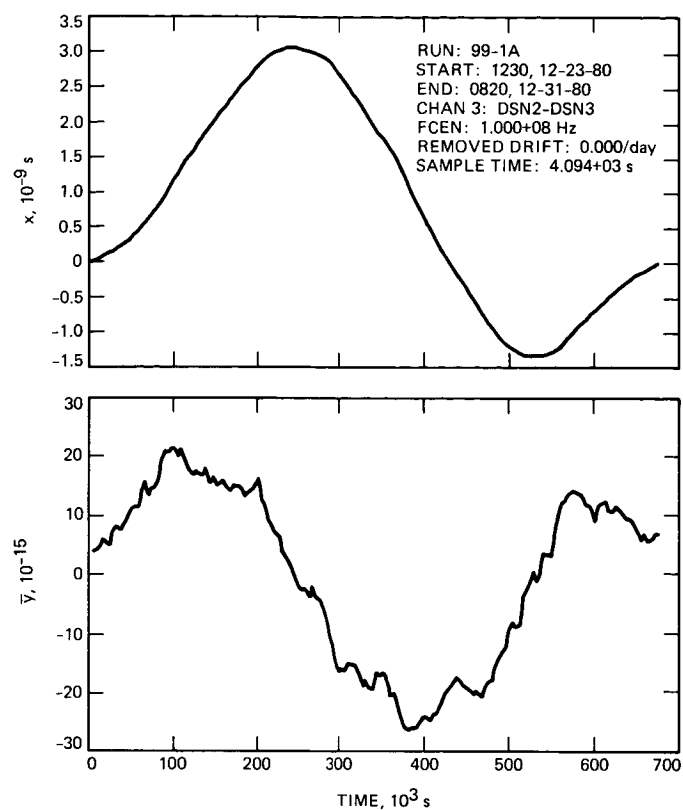
Fig. 1. x and ȳ from a test of hydrogen masers
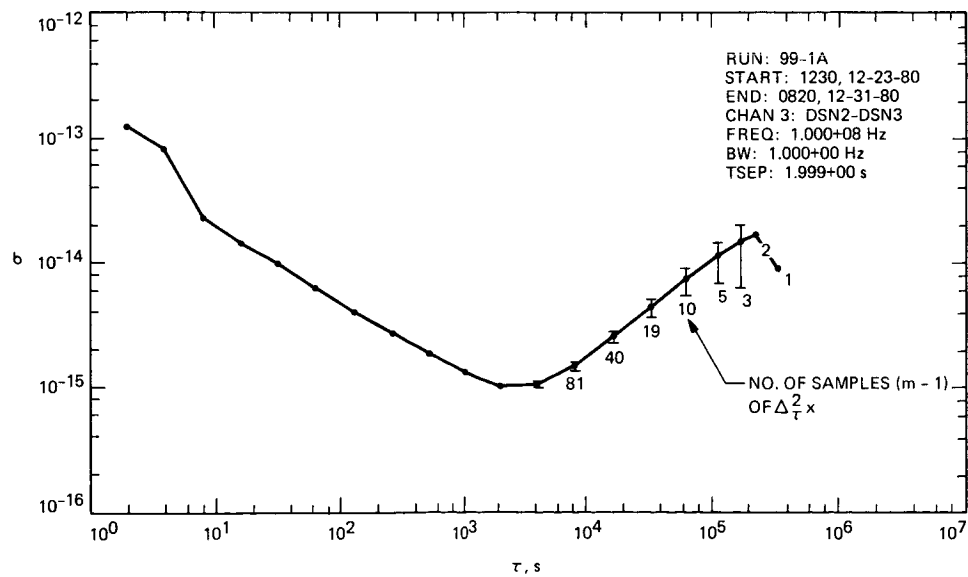
Fig. 2. Estimated Allan deviation of a pair of hydrogen masers
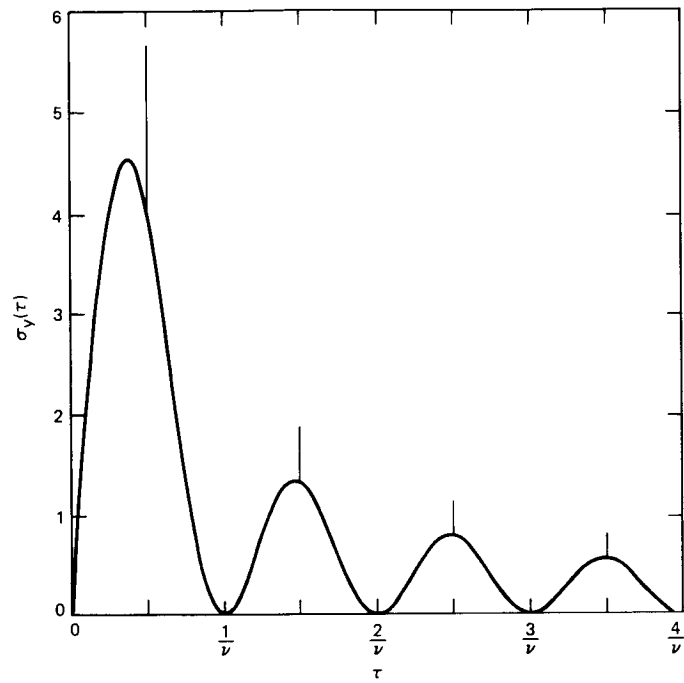
Fig. 3. Allan deviation for $x(t) = \cos(2\pi\nu t)$
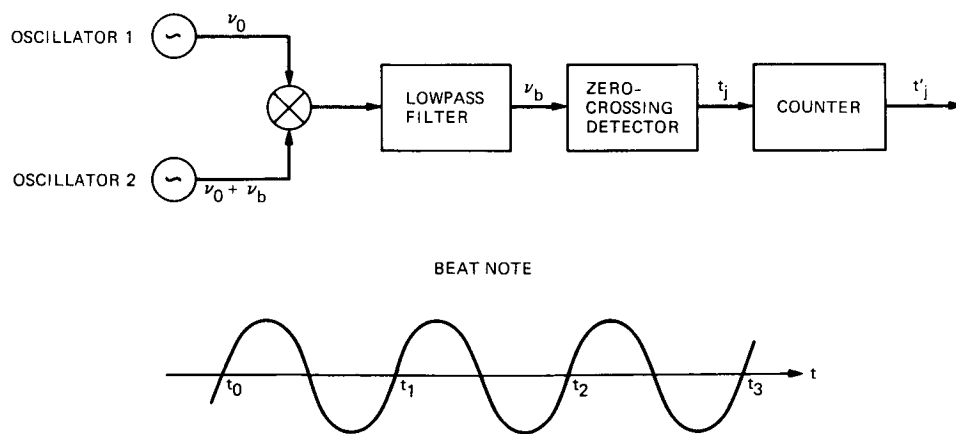
Fig. 4. Frequency stability test setup
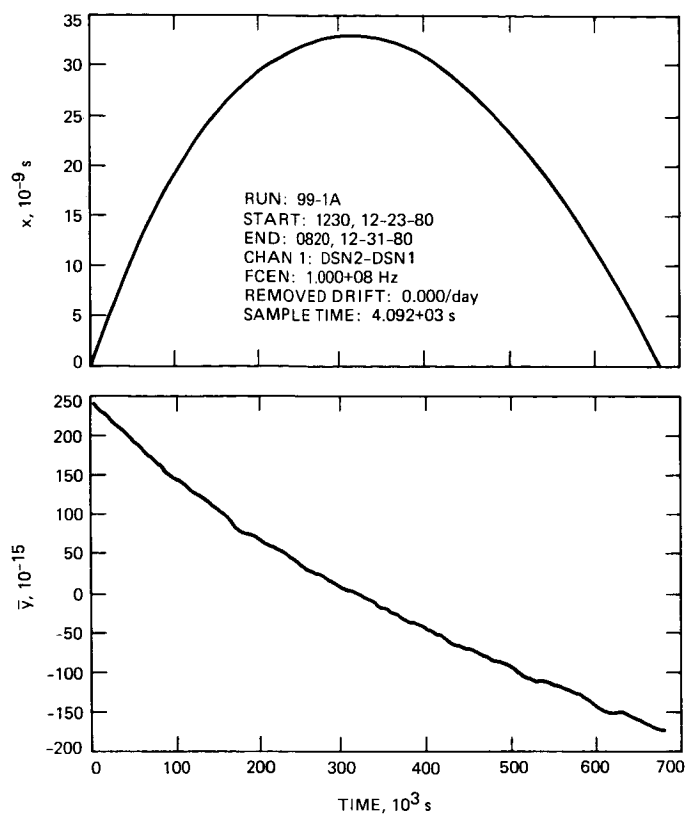
Fig. 5. x and ȳ before drift removal
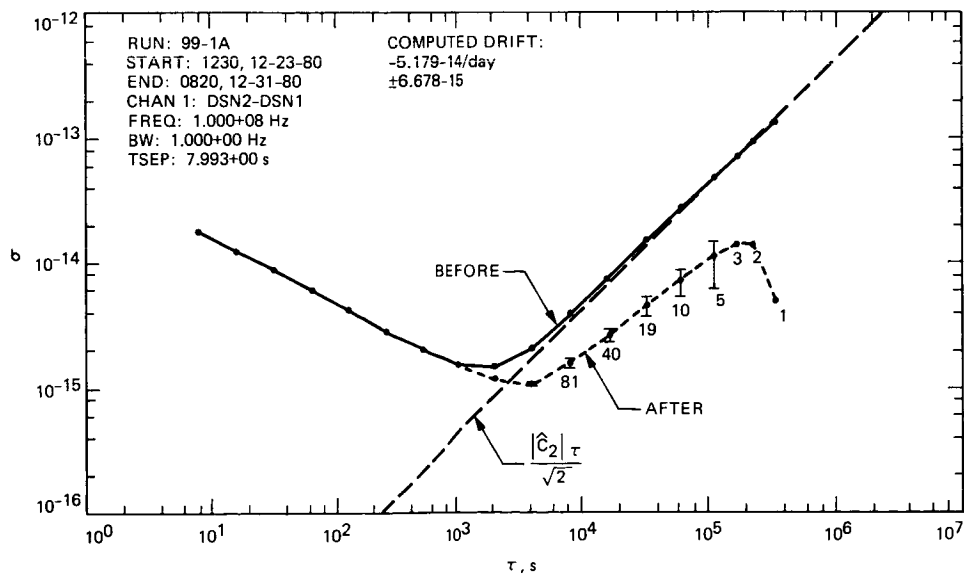
Fig. 6. Estimated Allan deviation before and after drift removal

RUN: 99-1A
START: 1230, 12-23-80
END: 0820, 12-31-80
CHAN 1: DSN2-DSN1
FCEN: 1.000+08 Hz
REMOVED DRIFT: -5.179-14/day
SAMPLE TIME: 4.092+03 s
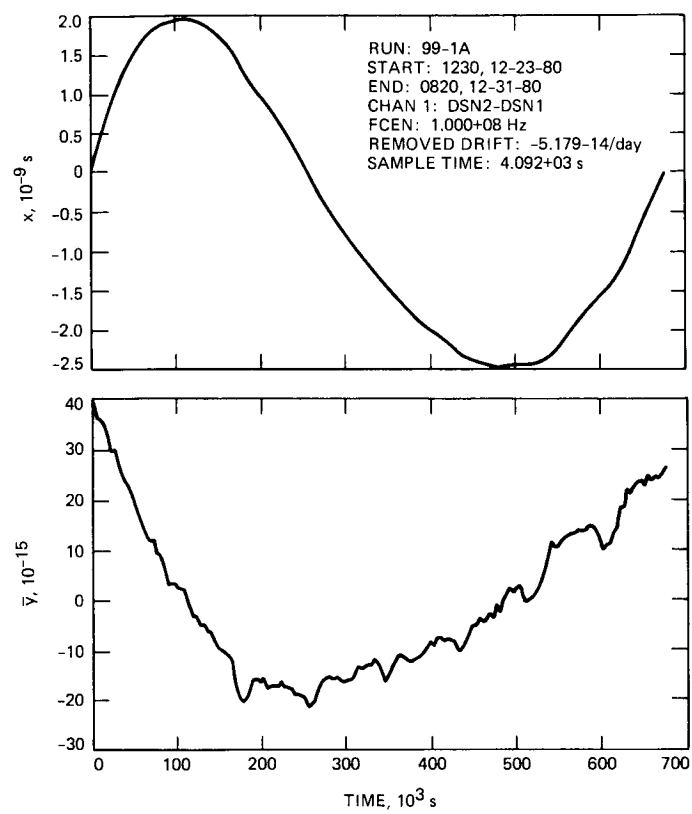
Fig. 7. x and $\bar{y}$ after drift removal
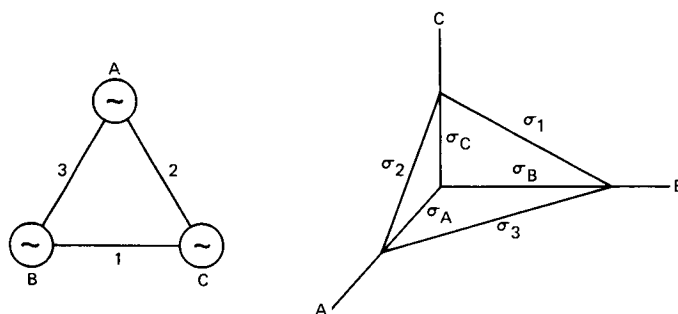
Fig. 8. Three-cornered hat

Network Functions and Facilities

R. J. Amorose

TDA Mission Support and DSN Operations Office

The objectives, functions, and organization of the Deep Space Network are summarized; tracking complexes, ground communications, and network operations control capabilities are described.

The Deep Space Network (DSN) was established by the National Aeronautics and Space Administration (NASA) Office of Space Tracking and Data Systems and is under the system management and technical direction of the Jet Propulsion Laboratory (JPL). The network is designed for two-way communications with (1) unmanned spacecraft traveling from earth to the planets and to the edge of our solar system, (2) unmanned satellites orbiting the earth, and (3) manned space shuttle missions. It has provided tracking and data acquisition support for a large number of deep space and earth orbiting exploration projects as listed in Table 1.

From its inception, NASA has had the objective of conducting scientific investigations throughout the solar system. It was recognized that in order to meet this objective, significant supporting research and advanced technology development must be conducted in order to provide deep space telecommunications for science data return in a cost effective manner. Therefore, the Network is continually evolving to keep pace with the state of the art of telecommunications and data handling. It was also recognized early that close coordination would be needed between the requirements of the flight projects for data return and the capabilities needed in the Network. This close collaboration is effected by the appointment of a Tracking and Data Systems Manager as part of the flight project team from the initiation of the project to the end of the mission. By this process, requirements are identified early enough to provide funding and implementation in time for use by the flight project in its flight phase.

The computer software necessary for the network operations control and monitor functions operates in computers that are configured in parallel to project data processing computers. A characteristic of the interface is that the Network provides direct data flow to and from the complexes, namely, metric data, command data science and engineering telemetry, and such network monitor data as are useful to the flight project. This is done via appropriate ground communication equipment to mission operations centers, wherever they may be.

The principal deliverables to the users of the Network are carried out by data system configurations as follows:

(1) The DSN Tracking System generates radio metric data, i.e., angles, one- and two-way doppler and range, and VLBI, and delivers magnetic tape data records to Mission Control.

(2) The DSN Telemetry System receives, decodes, records, and retransmits engineering and scientific data from the spacecraft to Mission Control via the Ground Communication Facility.

(3) The DSN Command System accepts spacecraft commands from Mission Control and transmits the commands via the Ground Communication Facility to a Deep Space Station. The commands are then radiated to the spacecraft in order to initiate spacecraft functions in flight.

(4) The DSN Radio Science System generates radio science data, i.e., the frequency and amplitude of spacecraft transmitted signals affected by passage through media such as the solar corona, planetary atmospheres, and planetary rings, and delivers magnetic tape data record to Mission Control.

The data system configurations supporting testing, training, and network operations control functions are as follows:

(1) The DSN Monitor and Control System instruments, transmits, records, and displays those parameters of the DSN necessary to verify configuration and validate the Network performance. It provides the tools necessary for Network Operations personnel to control and monitor the Network and interface with flight project mission control personnel.

(2) The DSN Test and Training System generates and controls simulated data to support development, test, training and fault isolation within the DSN. It participates in mission simulation with flight projects.

The capabilities needed to carry out the above functions have evolved in three technical areas:

(1) The Deep Space Complexes, which are distributed around the Earth at approximately 120° longitudinal separation in order to provide near continuous coverage of Deep Space missions. The technology involved in equipping these stations is strongly related to the state of the art of telecommunications and flight-ground design considerations, and is multimission in character.

(2) The Ground Communications Facility provides the capability required for the transmission, reception, and monitoring of Earth-based, point-to-point communications between the stations and the Network Operations Control Center at JPL, Pasadena, and to the JPL Mission Operations Centers. Four communications disciplines are provided: teletype, voice, high-speed, and wideband.

The Ground Communications Facility uses the capabilities provided by common carriers throughout the world, engineered into an integrated system by NASA Communications (NASCOM) which is managed by the Goddard Space Flight Center, and controlled from the Communications Center located in the Space Flight Operations Facility (Building 230) at JPL.

The Network Operations Control Center is the functional entity for centralized operational control of the Network and interfaces with the users. It has two separable functional elements: namely, Network Operations Control and Network Data Processing. The functions of the Network Operations Control are as follows:

Control and coordination of Network support to meet commitments to Network users.

Utilization of the Network data processing computing capability to generate all standards and limits required for Network operations.

Utilization of Network data processing computing capability to analyze and validate the performance of all Network systems.

The personnel who carry out the above functions are located in the Space Flight Operations Facility. Network personnel are directed by an Operations Control Chief. The functions of the Network Data Processing are as follows:

Processing of data used by Network Operations Control for control and analysis of the Network.

Display in the Network Operations Control Area of data processed in the Network Data Processing Area.

Interface with communications circuits for input to and output from the Network Data Processing Area.

Data logging and production of the intermediate data records.

The personnel who carry out these functions are also located in the Space Flight Operations Facility. The equipment consists of minicomputers for real-time data system monitoring, XDS Sigma 5s, VAX 11/780s, magnetic tape recorders, and appropriate interface equipment with ground data communications.

Table 1. Projects supported by the DSN

Mission	Project Management	Mission Operations
Ranger	JPL	JPL
Surveyor	JPL	JPL
Mariner Venus 1962	JPL	JPL
Mariner Mars 1964	JPL	JPL
Mariner Venus 1967	JPL	JPL
Mariner Mars 1969	JPL	JPL
Mariner Mars 1971	JPL	JPL
Mariner Venus–Mercury 1973	JPL	JPL
Voyager Jupiter–Saturn–Uranus	JPL	JPL
Lunar Orbiter	Langley Research Center	Langley Research Center
Pioneer	Ames Research Center	Ames Research Center
Apollo	Johnson Space Center	Johnson Space Center
Helios	German Spaceflight Operations Center	German Spaceflight Operations Center
Viking	Langley Research Center	Langley Research Center
Giotto	European Space Agency	European Space Agency
Sakigake	Institute of Space and Astronautical Science	Institute of Space and Astronautical Science
Suisei	Institute of Space and Astronautical Science	Institute of Space and Astronautical Science
International Cometary Explorer	Goddard Space Flight Center	Goddard Space Flight Center
Nimbus–7	Goddard Space Flight Center	Goddard Space Flight Center
Dynamic Explorer–1	Goddard Space Flight Center	Goddard Space Flight Center
International Sun–Earth Explorers 1 and 2	Goddard Space Flight Center	Goddard Space Flight Center
Shuttle	Johnson Space Center	Johnson Space Center

Author Index,[1] 1986

The Telecommunications and Data Acquisition Progress Report

42-85, January–March, May 15, 1986
42-86, April–June, August 15, 1986
42-87, July–September, November 15, 1986
42-88, October–December, February 15, 1987

Aguirre, S.

42-86   Phase Lock Acquisition for Sampled Data PLLs Using the Sweep Technique, pp. 95–102.

S. Aguirre, D. H. Brown, and W. J. Hurd

42-86   A Method to Dramatically Improve Subcarrier Tracking, pp. 103–110.

See W. J. Hurd.

42-87   A Comparison of Methods for DPLL Loop Filter Design, pp. 114–124.

S. Aguirre, W. J. Hurd, R. Kumar, and J. Statman

Ahlstrom, H. G.

42-85   Reverse Osmosis Water Purification System, pp. 181–190.

H. G. Ahlstrom, P. S. Hames, and F. J. Menninger

42-87   Antenna Servo Control System Characterization: Rate Loop Analysis for 34-m Antenna at DSS-15, pp. 165–173.

See Nickerson, J. A.

Amorose, R. J.

42-88   Network Functions and Facilities, pp. 213–215.

Anderson, T. O.

42-86   Antenna Arraying of Voyager Telemetry Signals by Symbol Stream Combining, pp. 131–142.

See Hurd, W. J.

Andres, E. M.

42-87   PPM/NAR 8.4-GHz Noise Temperature Statistics for DSN 64-Meter Antennas, 1982–1984, pp. 185–192.

See Slobin, S. D.

Armstrong, J. W.

42-85   Parkes-CDSCC Telemetry Array: Equipment Design, pp. 85–110.

See Brown, D. W.

Bathker, D. A.

42-86   A Prototype DSN X/S-Band Feed: Model III Development, pp. 26–32.

D. A. Bathker and H. F. Reilly

42-86   VLA Feedhorn for Voyager Encounter of Neptune, pp. 216–226.

See Manshadi, F.

42-87   Design Considerations for the Beam-Waveguide Retrofit of a Ground Antenna Station, pp. 10–23.

See Veruttipong, T.

---

[1] In the case of joint authorship, the reader is referred to the citation under the first author where all authors of the article are listed.